

Spoden, Christian; Leutner, Detlev

## Vergleichsarbeiten

2011, 41 S.



Quellenangabe/ Reference:

Spoden, Christian; Leutner, Detlev: Vergleichsarbeiten. 2011, 41 S. - URN:

urn:nbn:de:0111-pedocs-107492 - DOI: 10.25656/01:10749

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-107492>

<https://doi.org/10.25656/01:10749>

### Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

### Kontakt / Contact:

**peDOCS**

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung

E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)

Internet: [www.pedocs.de](http://www.pedocs.de)

Mitglied der

  
Leibniz-Gemeinschaft

---

# Vergleichsarbeiten

Christian Spoden

Detlev Leutner



---

## UDiKom

**Aus- und Fortbildung der Lehrkräfte  
in Hinblick auf Verbesserung der  
Diagnosefähigkeit, Umgang mit  
Heterogenität, individuelle Förderung**

Die Produktion dieses Materials  
zum Einsatz in die Lehrerbildung  
wurde ermöglicht durch

Deutsche Telekom Stiftung





## Vergleichsarbeiten

Alle im Projekt erstellten Materialien  
finden Sie unter  
[www.udikom.de](http://www.udikom.de)





## 2 Vergleichsarbeiten

<b>2.1 Gegenstand und Zielsetzungen</b>	3
2.1.1 Hintergrund	3
2.1.2 Gegenstand und Charakteristika von Vergleichsarbeiten	4
2.1.3 Allgemeine Zielsetzungen	5
2.1.3.1 Bestandsaufnahme fachlicher Kompetenzen und erreichter Lernstände, primär auf der Ebene der Klassen und Schulen	5
2.1.3.2 Didaktische und pädagogische Impulssetzung für eine datengestützte Unterrichtsentwicklung	6
2.1.3.3 Identifikation von Förderbedarf in Schülergruppen	6
2.1.3.4 Entwicklung und Stärkung der diagnostischen Kompetenz von Lehrkräften, insbesondere Stärkung der kriterialen Perspektive an Standards	6
2.1.4 Weiterführende Literatur	7
2.1.5 Verständnis und Diskussionspunkte	7
<b>2.2 Bewertungskriterien bei Vergleichsarbeiten</b>	8
2.2.1 Kriteriale Bezugsnorm bei Vergleichsarbeiten	8
2.2.2 Soziale Bezugsnorm bei Vergleichsarbeiten	9
2.2.2.1 Faire Vergleiche	9
2.2.3 Individuelle Bezugsnorm bei Vergleichsarbeiten	12
2.2.4 Weiterführende Literatur	12
2.2.5 Verständnis und Diskussionspunkte	12
<b>2.3 Testkonstruktion</b>	13
2.3.1 Die Beschreibung von Leistungsanforderungen durch aufeinander aufbauende Kompetenzstufen	13
2.3.2 Definition von Kompetenzstufen über Lösungswahrscheinlichkeiten: Das Rasch-Modell	14
2.3.3 Aufgabenauswahl bei Vergleichsarbeiten: Orientierung an Bildungsstandards und hohe Testgüte	15
2.3.4 Weiterführende Literatur	15
2.3.5 Verständnis und Diskussionspunkte	15
<b>2.4 Inhaltlicher Anwendungsbereich/Phänomenbereich</b>	16
2.4.1 Vergleichsarbeiten in der dritten Jahrgangsstufe (VERA 3)	16
2.4.1.1 Kompetenzskalen im Fach Deutsch bei VERA 3	16
2.4.1.2 Kompetenzskalen im Fach Mathematik bei VERA 3	19
2.4.2 Vergleichsarbeiten in der achten Jahrgangsstufe (VERA 8)	20
2.4.2.1 Kompetenzskalen im Fach Deutsch bei VERA 8	21
2.4.2.2 Kompetenzskalen im Fach Englisch bei VERA 8	23
2.4.2.3 Kompetenzskalen im Fach Mathematik bei VERA 8	26
2.4.3 Weiterführende Literatur	28
2.4.4 Verständnis und Diskussionspunkte	28
<b>2.5 Praktische Implikationen: Ergebnisnutzung aus Vergleichsarbeiten und Unterrichtsentwicklung</b>	29
2.5.1 Interpretation der Rückmeldungen aus Vergleichsarbeiten	29
2.5.2 Von der Ergebnisinterpretation zur schulischen Entwicklung	32
2.5.3 Ein Beispiel für die Ergebnisnutzung in den Fachkonferenzen	33
2.5.4 Kompetenzorientierter Unterricht	35
2.5.5 Weiterführende Literatur	39
2.5.6 Verständnis und Diskussionspunkte	39
<b>2.6 Literaturverzeichnis</b>	40

## 2.1 Gegenstand und Zielsetzungen

Vergleichsarbeiten stellen ein relativ junges Instrument der pädagogisch-psychologischen Diagnostik dar, um dessen Zielsetzungen es in der Lehrerschaft große Unsicherheit gibt. Im folgenden Kapitel wird zunächst der Entwicklungshintergrund von Vergleichsarbeiten vorgestellt. Im Anschluss werden dann der Gegenstandsbereich in Abgrenzung zur Individualdiagnostik und dem Bildungsmonitoring sowie die auf die Nutzung für schulinterne Evaluationszwecke ausgerichteten Zielsetzungen erläutert.

In diesem Kapitel werden folgende Fragen beantwortet:

- Wieso wurden Vergleichsarbeiten als dritte Säule der Qualitätssicherung in Deutschland implementiert?
- Mit welchen Zielsetzungen sind Vergleichsarbeiten ausgestattet und wie lassen sich diese von Zielsetzungen der Individualdiagnostik und dem Bildungsmonitoring abgrenzen?
- Durch welche charakteristischen Merkmale lassen sich Vergleichsarbeiten von Individualdiagnostik und Bildungsmonitoring abgrenzen?

### 2.1.1 Hintergrund

Das erwartungswidrig schwache Abschneiden deutscher Schülerinnen und Schüler im Vergleich der OECD-Staaten bei den internationalen Schulleistungsvergleichsstudien TIMSS II (Third International Mathematics and Science Study; Baumert et al., 1997) und PISA 2000 (Programme for International Student Assessment; Baumert et al., 2001; Baumert et al., 2002; vgl. Studienbrief 3) veranlasste die deutsche Kultusministerkonferenz (KMK) zur Vorbereitung weitreichender Veränderungen im deutschen Bildungssystem. Diese sollten die Qualitätssicherung von Unterricht und Schule auf Basis der *Setzung, Normierung und Überprüfung von Bildungsstandards* in den Vordergrund rücken. In Kooperation mit dem neu gegründeten „Institut zur Qualitätsentwicklung im Bildungswesen“ (IQB; <http://www.iqb.hu-berlin.de/>) formulierte die KMK daher eine Gesamtstrategie zum Bildungsmonitoring, die einen „Dreiklang aus mehr Eigenständigkeit für Schulen bei gleichzeitiger Vorgabe verbindlicher Standards und bei regelmäßiger Evaluation“ betonte (KMK & IQB, 2006). Als Reaktion auf den „PISA-Schock“ wurden durch die Bundesländer aber auch eigenständig zahlreiche Maßnahmen ergriffen, die zur Weiterentwicklung und Qualitätssicherung des Unterrichts in den Schulen führen sollten. Darunter lässt sich im Rahmen der *Standardsetzung* die an den nationalen Bildungsstandards orientierte und diese konkretisierende Überarbeitung der Lehrpläne fassen. Als Maßnahme der *Standardüberprüfung* führten nahezu alle Bundesländer außerdem flächendeckende Vergleichsarbeiten (in einigen Bundesländern unter den Namen *Diagnosearbeiten, Jahrgangsstufentest, Lernstandserhebungen, Kompetenztests oder Orientierungsarbeiten*) in der Grundschule und Sekundarstufe I ein. Die KMK griff nun wiederum ihrerseits diese Vorarbeiten der Länder auf und beschloss, regelmäßig durchgeführte Vergleichsarbeiten neben den internationalen Schulleistungsstudien, dem Vergleich der Bundesländer und der gemeinsamen Bildungsberichterstattung von Bund und Ländern in die oben skizzierte Gesamtstrategie einzubinden (KMK, 2006; vgl. auch Abb. 1).

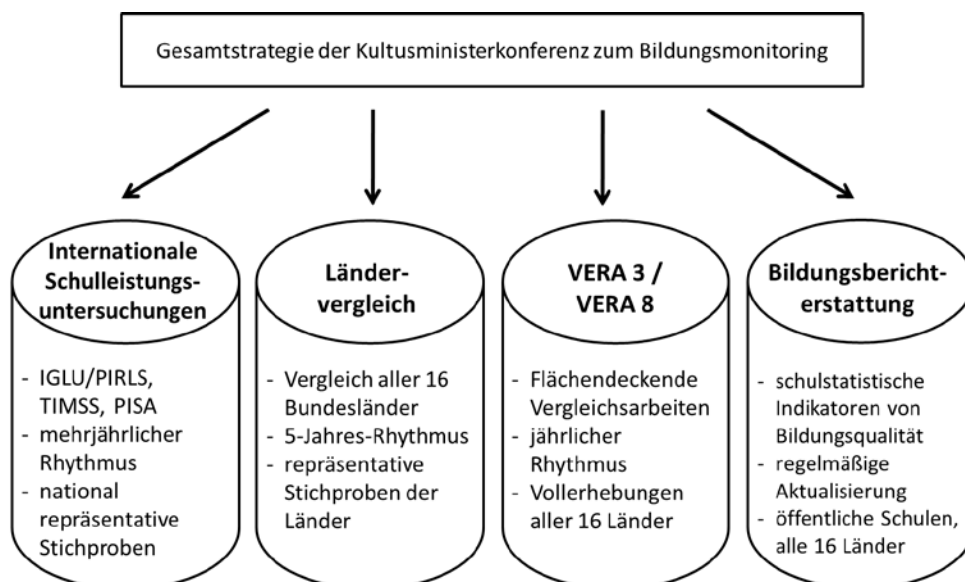


Abbildung 1: Vier Säulen der Qualitätssicherung in deutschen Schulen entsprechend der Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring (KMK, 2006)

Adressaten der diagnostischen Informationen aus Vergleichsarbeiten sind in vielen Bundesländern zunächst die Fachkonferenzen innerhalb der Schulen. Ihnen wird im Rahmen einer objektiven Leistungsrückmeldung Orientierung gegeben, inwieweit die erwarteten Standards innerhalb einer Klasse und innerhalb einer Jahrgangsstufe erreicht werden konnten. Diese Rückmeldung über Schülerleistungen ist mit der Erwartung verbunden, pädagogische, didaktische und ggf. curriculare Veränderungen in Gang zu setzen. Vergleichsarbeiten dienen der „landesweiten,

Hintergrund

Adressaten

jahrgangsbezogenen Untersuchung des Leistungsstands aller Schulen und Klassen“ (KMK & IQB, 2006, S. 21) im Hinblick auf die länderübergreifenden Bildungsstandards. Es verbleibt in der Verantwortung der Fachkonferenzen, im Anschluss an die Ergebnissrückmeldung aus den Arbeiten Konsequenzen für das Lehren und Lernen in der Schule abzuleiten (vgl. hierzu Kapitel 2.5). Hieraus erwächst die pädagogische Herausforderung der Vergleichsarbeiten für Lehrkräfte.

### 2.1.2 Gegenstand und Charakteristika von Vergleichsarbeiten

#### Merkmale von Vergleichsarbeiten

*Vergleichsarbeiten*<sup>1</sup> sind schriftliche Arbeiten der Schülerinnen und Schüler, die auf Basis vorgegebener Aufgabenstichproben landesweit in Teilleistungsbereichen ausgewählter Kernfächer mit dem Ziel durchgeführt werden, Schulleistungen orientiert an einer kriterialen und sozialen Bezugsnorm zu erfassen (vgl. Helmke & Hosenfeld, 2003a). Die Arbeiten finden in den Fächern Deutsch und Mathematik in der Jahrgangsstufe 3 sowie den Fächern Deutsch, Mathematik und der ersten Fremdsprache (Englisch/Französisch) in der Jahrgangsstufe 8 an einheitlichen Tagen und unter standardisierten Bedingungen statt.<sup>2</sup> Die Kompetenzerwartungen sind durch die nationalen Bildungsstandards für die Grundschule (KMK, 2004a-b) bzw. den mittleren Bildungsabschluss und Hauptschulabschluss (KMK, 2003a-c; 2004c-e) vorgegeben, welchen die Lehrpläne der Bundesländer verpflichtet sind. Aus den Bildungsstandards werden in jedem Jahr für die Vergleichsarbeiten bestimmte Kompetenz- oder Inhaltsbereiche ausgewählt.

Im Gegensatz zu Klassenarbeiten, welche das unmittelbar vorangegangene Unterrichtsgeschehen thematisieren, beziehen sich *Vergleichsarbeiten* auf die Erfassung von Kompetenzen (Fertigkeiten, Fähigkeiten, Kenntnisse; vgl. die folgende Erläuterung) bis zu einem festgelegten Zeitpunkt in der Bildungsbiographie. Als ein weiteres Merkmal von Vergleichsarbeiten sei die Einhaltung von Testgütekriterien genannt. Die Tests sind dazu im Vorfeld umfangreich erprobt worden (vgl. Kapitel 2.3.3; [http://www.iqb.hu-berlin.de/vera/wissrahmen?reg=r\\_4](http://www.iqb.hu-berlin.de/vera/wissrahmen?reg=r_4); 22.06.2011).

### Kompetenzen

Einer verbreiteten Definition von Weinert (2001) folgend, lassen sich Kompetenzen beschreiben als „die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können.“ Klieme & Leutner (2006) definieren in einer für die Bildungsforschung bedeutsamen Definition „Kompetenzen als kontextspezifische kognitive Leistungsdispositionen, die sich funktional auf Situationen und Anforderungen in bestimmten Domänen beziehen.“, betonen also die kognitive Komponente in Abgrenzung von motivationalen und emotionalen Aspekten. Gemeinsame Merkmale beider Definitionen sind die Orientierung an prinzipiell erfassbaren Leistungen (Fähigkeiten und Fertigkeiten) sowie der Kontextbezug (Kompetenzen sind *domänenspezifisch*). In Abgrenzung zu Intelligenz erweisen sich Kompetenzen außerdem in einem deutlich stärkeren Maße als lern- und trainierbar. Köller (2009) verweist auf ein pragmatisches Verständnis von Kompetenzen in den Bildungsstandards: Kompetenz ist hier durch das „gezeigte Verhalten“ in Bezug auf die Anforderungen der Standards definiert. Eine umfassende Diskussion des Kompetenzbegriffs in der Bildungsforschung nehmen Klieme, Hartig & Rauch (2008) vor.

#### Entwicklung der Tests, Durchführung und Ergebnisinterpretation

Die Erstellung der Testinstrumente wurde 2009 zentral durch das IQB übernommen, wobei die Projektgruppe VERA (<http://www.uni-landau.de/vera/>) für einige Bundesländer eine internetbasierte Ergebnissrückmeldung der Vergleichsarbeiten durchführt (vgl. Kapitel 2.4). Die Durchführung der Arbeiten und die Aufgabenbewertung erfolgen auf Basis standardisierter Manuale dezentral in den Schulen durch Lehrerinnen und Lehrer. Die Testauswertung und Ergebnissrückmeldung wird wiederum zentral durch die verantwortlichen Ministerien und Landesinstitute in Kooperation mit dem IQB (und gegebenenfalls dem VERA-Projektteam) vorgenommen. Die Ergebnisinterpretation liegt schließlich in den Händen der Fachkonferenzen in den Schulen, wobei diese mit umfangreichen Informationen, insbesondere ausführlichen didaktischen Kommentierungen, versorgt werden, welche die Interpretation erleichtern.

Bei Vergleichsarbeiten wird in den sprachbezogenen Inhaltsbereichen eine inhaltliche Schwerpunktsetzung auf einen oder mehrere Teilleistungs- oder Kompetenzbereiche mit dem Ziel vorgenommen, die Ausprägung der getesteten Kompetenzen in zu ermitteln (Tabelle 1, S. 5). Die Schwerpunktsetzung innerhalb eines Faches wechselt von Jahr zu Jahr, sodass die gesamte Breite dieses Faches im Laufe weniger Jahre über die verschiedenen Kompetenzbereiche hinweg abgebildet werden kann. Zum Einsatz kommen Aufgaben- bzw. Itemstichproben, welche repräsentativ für die jeweiligen Kompetenzerwartungen sind und zumeist einheitlich innerhalb eines Bildungsganges eingesetzt werden. Beabsichtigt ist es bei Vergleichsarbeiten nicht, eine möglichst reliable Schätzung der Kompetenzausprägung für eine einzelne Schülerin bzw. einen einzelnen Schüler zu erzielen (Individualdiagnostik; Studienbrief 1) oder aber an-

<sup>1</sup> Vergleichsarbeiten sollen aufgrund der begrifflichen Ähnlichkeit in einigen Bundesländern von *Parallelarbeiten* abgegrenzt werden. Bei diesen bearbeiten die Parallelklassen einer Schule dieselben Aufgabensätze, um so einen Leistungsvergleich innerhalb einer Schule über den Klassenverband hinaus zu ermöglichen. Diese sind in der hier vorgestellten Notation nicht mit *Vergleichsarbeiten* gemeint.

<sup>2</sup> In einigen Bundesländern erfolgen weitere Formen der Lernstandsdiagnose. Siehe hierzu Kapitel 2.4.

hand von Stichproben die Ausprägungen von Kompetenzen auf Landes- oder Bundesebene in der Breite eines Faches zu erfassen (Bildungsmonitoring; Studienbrief 3).

Abgrenzung  
von Individual-  
diagnostik /  
Bildungs-  
monitoring

	Individual- diagnostik	Vergleichsarbeiten	Bildungsmonitoring
<b>Zielsetzung</b>	Vorbereitung von Entscheidungen im Einzelfall	Vorbereitung pädagogischer, didaktischer und/ oder curricularer Entscheidungen auf Schul- und Unterrichtsebene (Selbstevaluation)	Vorbereitung politischer Entscheidungen auf Schulsystemebene (Fremdevaluation)
<b>Fokus</b>	<p>Inhaltliche Tiefe in einem Fachgebiet:</p> <p>→ Aufgaben bzw. Itemstichproben</p> <p>→ Schätzung der Ausprägung einer Eigenschaft einer Schülerin/eines Schülers</p>	<p>Zunächst Tiefe in einem Fachgebiet, ggf. dann Breite durch Abdeckung weiterer Fachgebiete eines Faches in den Folgejahren:</p> <p>→ Aufgaben bzw. Itemstichproben</p> <p>→ Kompetenzerhebung der Schülerinnen und Schüler in Klassen und Jahrgangsstufen (keine Individualdiagnostik)</p>	<p>Fachliche Tiefe und zugleich fachliche Breite:</p> <p>→ Aufgaben bzw. Itemstichproben und Stichprobe der Schülerinnen und Schüler (Multiple-Matrix-Stichprobe; vgl. Studienbrief 3)</p> <p>→ Kompetenzerhebung in Bundesländern, Staaten (keine Individualdiagnostik, keine Aussagen auf Schul- und Klassenebene)</p>

Tabelle 1: Zielsetzung von Individualdiagnostik, Vergleichsarbeiten und Bildungsmonitoring

In Kapitel 2.4 wird erläutert, wie sich die aktuelle Konzeption von Vergleichsarbeiten in den jeweiligen Unterrichtsfächern darstellt.

### 2.1.3 Allgemeine Zielsetzungen

Die Bundesländer haben in Bezug auf Vergleichsarbeiten eine unterschiedlich lange Tradition mit teilweise unterschiedlicher Gewichtung der Zielsetzungen. Übereinstimmend lassen sich aber folgende allgemeine Zielsetzungen von Vergleichsarbeiten festhalten (vgl. hierzu auch das Arbeitspapier der *Arbeitsgruppe Empirische Schulentwicklung, EMSE*, aus dem Jahr 2006):

allgemeine  
Zielsetzungen

- *Bestandsaufnahme fachlicher Kompetenzen und erreichter Lernstände, primär auf der Ebene der Klassen und Schulen*
- *didaktische und pädagogische Impulssetzung für eine datengestützte Unterrichtsentwicklung*
- *Identifikation von Förderbedarf in Lerngruppen*
- *Entwicklung und Stärkung der diagnostischen Kompetenz von Lehrkräften, insbesondere Stärkung der kriterialen Perspektive orientiert an Standards*

#### 2.1.3.1 Bestandsaufnahme fachlicher Kompetenzen und erreichter Lernstände, primär auf der Ebene der Klassen und Schulen

Bei Vergleichsarbeiten wird die Ausprägung von Kompetenzen ermittelt. Die Rückmeldung der Ergebnisse aus Vergleichsarbeiten umfasst nicht nur rein numerische Aussagen zum Ergebnis der Lerngruppe. Vielmehr sind die Rückmeldungen so aufbereitet, dass sie kriteriale Vergleiche (vgl. Studienbrief 1; siehe 2.2.1) im Hinblick auf das Erreichen bestimmter Leistungsanforderungen ermöglichen. Bei der Rückmeldung als Verteilung von Schülerinnen und Schülern auf *Kompetenzstufen*, werden diese Kompetenzstufen durch konkrete Aufgabenbeispiele inhaltlich so beschrieben, dass eine klare Vorstellung davon vermittelt werden kann, welchen Anforderungen die Schülerinnen und Schüler gerecht werden. Die Anforderungen der Testaufgaben sind durch die nationalen Bildungsstandards definiert und geben damit eine objektive Rückmeldung über das Erreichen der dort festgeschriebenen Anforderungen des jeweiligen Unterrichtsfaches. Vergleichsarbeiten ermöglichen den Schulen damit, im Vorfeld des Abschlusses der Primarstufe bzw. der Sekundarstufe I, eine schulinterne Standardüberprüfung.

Bestands-  
aufnahme  
fachlicher  
Kompetenzen  
und erreichter  
Lernstände

#### 2.1.3.2 Didaktische und pädagogische Impulssetzung für eine datengestützte Unterrichtsentwicklung

Flächendeckend durchgeführte Vergleichsarbeiten haben die Funktion, pädagogische, didaktische und curriculare Entscheidungen auf der Ebene der Einzelschule anzustoßen. Im Rahmen der Ergebnisinterpretation müssen in den Schulen, zumeist innerhalb der Fachkonferenzen, die Ursachen für einen hohen oder niedrigen Ausprägungsgrad der getesteten Kompetenzen erkundet werden. Idealerweise werden in diesem Prozess Konsequenzen und Maßnahmen diskutiert, wie der Kompetenzerwerb der Schülerinnen und Schüler besser gefördert bzw. ein hohes Leistungsniveau

didaktische/  
pädagogische  
Impuls-  
setzung

auf lange Zeit gehalten werden kann. Differenzen in den erreichten Anforderungen zwischen den Fächern, zwischen Teilleistungsbereichen innerhalb eines Faches, zwischen den Klassen derselben Jahrgangsstufe einer Schule oder im Vergleich zu entsprechenden Referenzgruppen können als Anknüpfungspunkte für inhaltliche Schwerpunktsetzungen dienen. Mit diesen Erwartungen ausgestattet, entsprechen Vergleichsarbeiten tendenziell eher dem Konzept der Selbstevaluation (siehe Erläuterungen unten), da zwar Testverfahren zentral erstellt, diese aber dezentral in den Schulen von den Lehrkräften selbst eingesetzt, kodiert und bezüglich ihrer Ergebnisse interpretiert werden.

### Selbst- vs. Fremdevaluation im Bildungswesen

Ruep und Keller (2007) beschreiben das Verhältnis von Selbst- und Fremdevaluation als „Vergleich zweier Wahrnehmungen“. Bei der Selbstevaluation findet eine Bewertung pädagogischer Arbeit nicht durch externe Experten statt, sondern durch die verantwortlichen Personen vor Ort (in der Regel also die Schulleitung, die didaktischen Leitung, Koordinatorinnen und Koordinatoren, Fachkonferenzen). Die Bezeichnung Selbstevaluation beinhaltet allerdings nicht, dass der gesamte Evaluationsprozess von verantwortlichen Personen der Schulen übernommen werden muss; jedoch bleibt die Schule selbst verantwortlich für die spezifischen Ziele der Evaluation, die Verwendung und Interpretation ihrer Ergebnisse sowie die Auswahl und Umsetzung von Interventionsmaßnahmen. Die umgesetzten Maßnahmen bedürfen wiederum einer Bewertung, sodass sich häufig ein Zyklus aus Evaluations- und Interventions-schritten ergibt. Ziel der Selbstevaluation ist die Feststellung der Schul- und Unterrichtsqualität. Zur Fragestellung kann aber auch die optimale Nutzung vorhandener Ressourcen (Sach- und Personalmittel) werden. Die Ergebnisse der Selbstevaluation können außerdem zum Ausgangspunkt fremdevaluierender Maßnahmen werden oder aber diese ergänzen. Bei der *Fremdevaluation* entsteht durch die Bestandsaufnahme kritischer, externer Evaluatoren (in der Regel Mitglieder der Schulaufsicht oder anderer Schulen) ein objektives Bild des Lehrens und Lernens. Charakteristisch ist deren Bewertung von gut operationalisierten Indikatoren der Schul- und Unterrichtsqualität mit Hilfe standardisierter Verfahren der Beobachtung und Befragung. Ziele und Zeitpunkt der Evaluation werden mit der Schulleitung im Vorhinein abgeklärt. Die Schulleitung und das Kollegium werden nach Auswertung quantitativer und qualitativer Daten durch einen detaillierten Evaluationsbericht über die Ergebnisse unterrichtet und sollten im Anschluss die Möglichkeit zur Rückmeldung erhalten. Der Evaluationsbericht geht neben der Schule auch der Schulaufsicht zu, die weitere Maßnahmen der Qualitätssicherung mit der Schule diskutiert.

#### 2.1.3.3 Identifikation von Förderbedarf in Lerngruppen

Identifikation  
von Förder-  
bedarf in  
Schüler-  
gruppen

Kriteriale Vergleichsmaßstäbe ermöglichen die Identifikation von Schülergruppen, welche die erwarteten Standards verfehlen. Ergebnisse des Bildungsmonitorings (vgl. Studienbrief 3) haben verdeutlicht, dass die Leistungsstreuung in den Ländern der Bundesrepublik Deutschland hoch ist. Legt man die Ergebnisse der PISA-Erhebung 2006 zugrunde, so erreichen ca. 20 % der 15-jährigen Schülerinnen und Schüler in Deutschland in den Kompetenzbereichen Lesen und Mathematik nicht die Kompetenzstufe 2 und weisen damit besonders dringenden Förderbedarf auf. Es steht zu befürchten, dass diese Schülerinnen und Schüler erhebliche Schwierigkeiten haben werden, beruflichen und allgemein lebensrelevanten Anforderungen gerecht zu werden (Artelt, Stanat, Schneider & Schiefele, 2001). Diese Anteile von Schülerinnen und Schülern mit erheblichen Defiziten werden von Lehrkräften tendenziell unterschätzt (siehe Abschnitt 2.1.3.4), was möglicherweise auch damit zu tun hat, dass ihnen objektive Rückmeldungen über die Leistungsfähigkeit nicht in ausreichendem Maße zur Verfügung stehen. Alltag der meisten Lehrkräfte sind klasseninterne Vergleichsmaßstäbe, welche aber den tatsächlichen Förderbedarf von Schülerinnen und Schülern bei einer entsprechend leistungsschwachen Referenzgruppe (Klasse/Kurs) verbergen können. Vergleichsarbeiten ermöglichen hingegen, den Anteil förderungsbedürftiger Schülerinnen und Schüler vor dem Hintergrund der in den Bildungsstandards definierten Kompetenzerwartungen zu bestimmen.

#### 2.1.3.4 Entwicklung und Stärkung der diagnostischen Kompetenz von Lehrkräften, insbesondere Stärkung der kriterialen Perspektive orientiert im Hinblick auf Standards

Entwicklung  
und Stärkung  
diagnosti-  
scher  
Kompetenz

Diagnostische Kompetenzen von Lehrerinnen und Lehrern werden zumeist als Fähigkeiten verstanden, Schülerinnen und Schüler hinsichtlich der Ausprägung bestimmter Merkmale treffend zu beurteilen (vgl. Schrader, 2001; Spinath, 2005). Des Weiteren lässt sich die Fähigkeit der Lehrkraft, Aufgaben hinsichtlich ihrer Schwierigkeit für die jeweilige Lerngruppe angemessen einzuschätzen, unter diagnostischer Kompetenz einordnen. Referenzrahmen ist hier die tatsächliche (empirische) Schwierigkeit der Aufgabe, in der Regel also die Lösungsquote.

Reflexion  
diagnostischer  
Überlegungen

Wie aber können Vergleichsarbeiten dazu beitragen, die diagnostischen Kompetenzen von Lehrerinnen und Lehrern zu stärken? Stellt man die treffende Beurteilung von Schülerinnen und Schülern in den Vordergrund, so erlauben die mit den Ergebnissrückmeldung verbundenen sozialen Vergleichsmöglichkeiten (vgl. Kapitel 2.2.2) eine objektive Leistungseinschätzung, unabhängig vom zuvor beschriebenen klasseninternen (bzw. schulinternen) Vergleichsmaßstab. Lehrkräften bietet sich die Möglichkeit, eigene Leistungserwartungen anhand der objektiven Ergebnisse von Referenzgruppen zu korrigieren. Mehr noch steht aber die Stärkung der *kriterialen* Perspektive der Lehrkräfte im Vordergrund (vgl. Kapitel 2.2): Von ihnen wird eine möglichst treffende Verortung des Leistungsstandes der Lerngruppe



in Bezug auf die nationalen Bildungsstandards erwartet, die bei Vergleichsarbeiten durch Überprüfung des tatsächlichen Lernstandes (z.B. über die Verteilung auf Kompetenzstufen) kritisch reflektiert werden kann.

Auch in Bezug auf den zweiten Aspekt diagnostischer Kompetenzen, die angemessene Einschätzung von Aufgaben, bieten die Vergleichsarbeiten einen Referenzmaßstab: Die Aufgabenbeispiele der Vergleichsarbeiten konkretisieren die Inhalte der Bildungsstandards und ermöglichen Lehrkräften somit ein besseres Verständnis dieser abstrakt formulierten Inhalte. Grundlage für die objektive Einschätzung der Schwierigkeit ist die z.B. im Jahr 2011 erfolgte Rückmeldung von Lösungsquoten für die Aufgaben in verschiedenen Bildungsgängen. Die Analyse der Testaufgaben bietet so in Bezug auf Inhalte und Schwierigkeitsgrad Anhaltspunkte für die Ableitung von Lernaufgaben, die kritische Gegenüberstellung vorhandener Unterrichtsmaterialien und die Ausarbeitung von kompetenzorientierter Unterrichtsreihen (vgl. Kapitel 2.5.4).

#### 2.1.4 Weiterführende Literatur

Entwicklung und Zielsetzungen von Vergleichsarbeiten sind durch die Länder auf ihren Internetseiten dargelegt Literatur worden (siehe auch [http://www.iqb.hu-berlin.de/vera?reg=r\\_1](http://www.iqb.hu-berlin.de/vera?reg=r_1)). Als repräsentative Auswahl der Literatur zum wissenschaftlichen Hintergrund können Helmke und Hosenfeld (2003a, b), Leutner, Fleischer, Spoden und Wirth (2007), Nachtigall und Jantowski (2004) sowie Peek (2004) empfohlen werden.

#### 2.1.5 Verständnis und Diskussionspunkte

1. *Nennen Sie Charakteristika von Vergleichsarbeiten.*
2. *Fassen Sie allgemeine Zielsetzungen von Vergleichsarbeiten zusammen. Versuchen Sie insbesondere zu skizzieren, wie diese Zielsetzungen ineinander greifen.*
3. *Legen Sie dar, wieso diese Zielsetzungen nicht durch bestehende Instrumente erfüllt werden konnten:*
  - *Klassenarbeiten*
  - *Parallelarbeiten*
  - *Instrumente des Bildungsmonitorings (IGLU, PISA, etc.).*

## 2.2 Bewertungskriterien bei Vergleichsarbeiten

Die Rückmeldung des absoluten Ergebnisses einer Klasse oder Jahrgangsstufe ist für sich betrachtet häufig wenig aussagekräftig und birgt die Notwendigkeit in sich, geeignete Maßstäbe für die Interpretation der Testergebnisse heranziehen zu können. In Studienbrief 1 sind in diesem Zusammenhang kriteriale, soziale und individuelle Bezugsnormen als Interpretationsmaßstäbe eingeführt worden. Im Folgenden wird verdeutlicht, welche Informationen als Bewertungskriterien zur Interpretation von Ergebnissrückmeldungen aus Vergleichsarbeiten genutzt werden können.

In diesem Kapitel werden folgende Fragen beantwortet:

- *Inwiefern stellen die nationalen Bildungsstandards eine kriteriale Bezugsnorm für die Ergebnisse aus Vergleichsarbeiten dar?*
- *Was ist unter Fairness sozialer Vergleiche zu verstehen und wie wird sichergestellt, dass soziale Vergleiche von Testergebnissen im Kontext von Vergleichsarbeiten fair gestaltet sind?*

### 2.2.1 Kriteriale Bezugsnorm bei Vergleichsarbeiten

Anforderungen  
der Bildungs-  
standards  
als Kriterium

Kriteriale Bezugsnormorientierung beschreibt die Interpretation von Testergebnissen in Bezug auf ein inhaltlich definiertes Kriterium. Die Inhalte der Testaufgaben von Vergleichsarbeiten sind durch die nationalen Bildungsstandards (KMK, 2003a-c; 2004a-e) festgelegt, welchen die landespezifischen Curricula verpflichtet sind. In den Bildungsstandards werden fachspezifisch Schülerkompetenzen festgelegt, die bis zu einem bestimmten Zeitpunkt in der Bildungsbiographie erworben werden sollen. Vergleichsarbeiten ermöglichen die schulinterne Überprüfung dieses Kompetenzerwerbs. Die Testaufgaben der Vergleichsarbeiten werden aus fachdidaktischer, pädagogischer und lehrerpsychologischer Sicht bezüglich ihrer Anforderungen analysiert. Als Testaufgaben kommen dabei nur solche Aufgaben in Betracht, deren Lösungsanforderungen die in den Bildungsstandards definierten Kompetenzen voraussetzen (vgl. das folgende Beispiel). Aktuell werden in den meisten Kompetenzstufenmodellen Aufgaben mit ähnlichem, inhaltlich beschreibbarem Anforderungsprofil im Rahmen der Testkonstruktion zusammengefasst und unter Berücksichtigung ihrer Schwierigkeit als Kompetenzstufen ausgewiesen (vgl. Kapitel 2.3).

Von der KMK sind so genannte Mindeststandards formuliert worden, die von allen Schülerinnen und Schülern erreicht werden sollen (vgl. Kap. 2.4). Die Erfüllung der Mindeststandards stellt ein zentrales Kriterium zur Ergebnisinterpretation bei Vergleichsarbeiten dar.

#### Beispiel:

Aufgaben-  
beispiele

Die nachfolgend dargestellten Aufgaben (Abbildung 2, S. 9, Abbildung 3, S. 10) wurden in den nationalen Bildungsstandards als Beispielaufgaben<sup>3</sup> für den Kompetenzbereich "Leseverstehen" im Fach Englisch veröffentlicht (KMK, 2004d). An ihnen soll die Validität der Aufgaben in Bezug auf die Bildungsstandards verdeutlicht werden.

Das Aufgabenbeispiel 1 beinhaltet einen (diskontinuierlichen) Sachtext mit 211 Wörtern und enthält zehn Items des Itemformats „true/false“. Die Aufgabe erfasst Kompetenzen, die noch bis zum Hauptschulabschluss im Fach Englisch erwartet werden.

Aufgabenbeispiel 3 umfasst einen fiktionalen Text mit 196 Wörtern. Sieben Items des Formats „multiple-choice“ erfassen ebenfalls Kompetenzen, die überwiegend bis zum Hauptschulabschluss beherrscht werden sollen.

In den Bildungsstandards werden die den Aufgaben zugrunde liegenden Anforderungsmerkmale folgendermaßen definiert (relevante Stellen jeweils fett markiert):

„Die Schülerinnen und Schüler können **kurze, einfache Texte lesen und verstehen**, die einen sehr **frequenten Wortschatz** und einen **gewissen Anteil international bekannter Wörter** enthalten (A 2).

Die Schülerinnen und Schüler können

- *kurze, einfache persönliche Briefe und E-Mails verstehen (A2),*
- ***konkrete, voraussagbare Informationen in einfachen Alltagstexten auffinden**, z.B. in **Anzeigen, Prospekten**, Speisekarten, Fahrplänen, Programmzeitschriften (A2),*
- *gebräuchliche Zeichen und Schilder an öffentlichen Orten, z.B. Wegweiser, Warnungen vor Gefahr verstehen (A2),*
- ***aus einfacheren schriftlichen Materialien** wie Briefen, Broschüren, Zeitungsartikeln (oder auch **dem Niveau entsprechenden fiktionalen Texten**) **spezifische Informationen herausfinden** (A2),*
- *einfache Anleitungen für Apparate verstehen, mit denen sie im Alltag zu tun haben (A2).“*

<sup>3</sup> Das Layout der Aufgaben wurde für den Zweck der besseren Darstellung angepasst.

The Scottish Seabird Centre, The Harbour, North  
Berwick, EH39 4SS  
tel: 01620 890 202  
email: info@seabird.org  
Registered Charity SCO25837

#### Open all year

Summer 10.00 – 18.00 Winter 10.00 – 16.00  
(weekdays)

10.00 – 17.30 (weekends)

Open all year, except Christmas Day

Last admission 45 minutes before closing

#### Admission charges

Adult: £4.95

Senior/Child Concession: £3.50

Family (2 adults + 2 children): £13.50

**Children under 5 get in free!**

Easy to get to

#### By Train:

Just 30 minutes from Edinburgh by a direct and frequent service. Special all-inclusive excursion Scotrail excursion fare only £7.50 for an adult ticket. Contact 08457 484950. The Centre is a pleasant 10 minute walk from the station (just follow the signs).

#### By Car

Only 25 miles/ 40kms from the centre of Edinburgh. The Scottish Seabird Centre and North Berwick are signposted from the A1.

#### By Bus

Regular bus services operate from Edinburgh (service no. 124/X5 run by First Edinburgh 0131 663 9233) from Haddington (no. 121 run by First Edinburgh) and from Dunbar (no. 120 run by Eve Coaches 01368 863455).

Don't miss the amazing spectacle of **grey seals with their fluffy white newborn pups** on the Isle of May National Nature Reserve from October to December, one of the largest grey seal colonies on the east coast of Great Britain.

#### Task:

- Read the information about the Scottish Seabird Centre at North Berwick on the East coast of Scotland.
- Read the following statements. Decide if they are "true" or "false" according to the text.

	true	false
1. You can visit the seabird Centre every day of the year.	<input type="checkbox"/>	<input type="checkbox"/>
2. On Fridays the Centre always closes at 4 pm.	<input type="checkbox"/>	<input type="checkbox"/>
3. A family ticket costs £ 14.95.	<input type="checkbox"/>	<input type="checkbox"/>
4. Children over 5 and senior citizens pay the same money.	<input type="checkbox"/>	<input type="checkbox"/>
5. You can only get information about the Centre by e-mail.	<input type="checkbox"/>	<input type="checkbox"/>
6. School classes can only go to the Centre by bus.	<input type="checkbox"/>	<input type="checkbox"/>
7. The Seabird Centre is in Haddington.	<input type="checkbox"/>	<input type="checkbox"/>
8. You can see seals with their young on the Isle of May.	<input type="checkbox"/>	<input type="checkbox"/>
9. It takes 30 minutes to walk to the Centre from the station.	<input type="checkbox"/>	<input type="checkbox"/>
10. Nobody is admitted thirty minutes before closing.	<input type="checkbox"/>	<input type="checkbox"/>

Abbildung 2: Aufgabenbeispiel 1 der Bildungsstandards für den Kompetenzbereich Leseverstehen im Fach Englisch in Anlehnung an KMK (2004d, S. 26 f.)

### 2.2.2 Soziale Bezugsnorm bei Vergleichsarbeiten

Soziale Vergleichsmaßstäbe für Lehrerinnen und Lehrer beschränken sich häufig auf klassen- oder jahrgangsstufen-interne Leistungsvergleiche und vernachlässigen schulübergreifende Vergleichsmöglichkeiten (Schrader & Helmke, 2001). Um die Leistungsfähigkeit einer Lerngruppe allerdings realistisch einschätzen zu können, ist es notwendig, ihren Lernstand mit Lerngruppen anderer Schulen vergleichen zu können. Für diesen sozialen Vergleich werden bei Vergleichsarbeiten Klassen oder Kurse ausgewählt, die sich aufgrund ihrer Ähnlichkeit in leistungsrelevanten Merkmalen als Referenzgruppe besonders eignen (vgl. Leutner, 2010). Vergleichsarbeiten sind zwar nicht vollends um die Strenge oder Milde im Urteil der Lehrkraft bereinigt (vgl. Leutner, Fleischer, Spoden & Wirth, 2007; vgl. Kapitel 2.3), Unterschiede in der persönlichen Urteilsstrenge zwischen Lehrerinnen und Lehrern werden aber aufgrund desselben Aufgabenmaterials in allen Referenzklassen, der überwiegenden Verwendung geschlossener Aufgabenformate und einer standardisierten Durchführungs- und Auswertungsanleitung minimiert. Damit soziale Vergleiche aber objektive Aussagen über die Qualität der Beschulung zulassen, ist es zudem notwendig, Schulen faire Referenzwerte bereitzustellen.

soziale  
Bezugsnorm

#### 2.2.2.1 Faire Vergleiche

Schulische Lerngelegenheiten stellen notwendige Voraussetzungen für den Aufbau von Schülerkompetenzen dar und die Qualität dieser Lerngelegenheiten hat einen entscheidenden Einfluss auf die Ausprägung der erworbenen Kompetenzen (Köller & Baumert, 2002; Helmke & Schrader, 2001). Darüber hinaus besitzen jedoch auch Merkmale der Schülerinnen und Schüler, die sich dem Einflussbereich von Schule und Unterricht entziehen, Einfluss auf den Aufbau von Kompetenzen. Darunter sind zum einen individuelle kognitive Eingangsvoraussetzungen, insbesondere Intelligenz, zu fassen (Helmke & Schrader, 2001; Süß, 2001). Doch auch sozioökonomische Merkmale der Schülerschaft, z.B. der elterliche Bildungsgrad und die Familiensprache, besitzen nachweislich einen Einfluss auf das schulische Leistungsniveau (Ramm, Prenzel, Heidemeier & Walter, 2004; Nachtigall, Kröhne, Enders & Steyer, 2008). Ergebnisse der PISA-2000-Studie haben verdeutlicht, dass die Unterschiede in der Zusammensetzung der Schülerschaft in Bezug auf Kontextmerkmale sogar innerhalb derselben Schulform beträchtlich sein können (vgl. Baumert et al., 2003). Soziale Vergleiche einer Klasse oder Jahrgangsstufe mit Vergleichspopulationen sollten dahingehend fair gestaltet sein, als dass leistungsrelevante Kontextmerkmale der Schülerschaft beim sozialen Vergleich Berücksichtigung finden

faire  
Vergleiche



müssen. Andernfalls bestünde die Gefahr, dass an Schulen mit ungünstigen Eingangsvoraussetzungen die Effektivität der schulischen Arbeit unterschätzt, an Schulen mit günstigen Eingangsvoraussetzungen überschätzt wird (Peek & Döbelstein, 2006). Das Ziel der Berücksichtigung von Kontext- oder Hintergrundinformationen ist daher, die tatsächliche Qualität der Beschulung von Leistungseinflüssen zu trennen, die aufgrund von Merkmalen der sozialen Herkunft oder individueller kognitiver Fähigkeiten zustande kommen.

<p><b>Ingrid McFarlane</b> <b>Zoo Keeper</b></p> <p>When I left school at eighteen, I got a job at a zoo as a student keeper. Now, five years later, things have changed – I have passed my exams and I am a full animal keeper.</p> <p>The money is not good. I only get £9,000 a year. You have to be outside in rain and snow, which is hard work, and you get very dirty. But this doesn't matter to me because animals are the most important thing in my life!</p> <p>There are a hundred monkeys and fifty deer in my part of the zoo and I give them their food and clean their houses. I also need to watch them carefully to be sure that they are all well. In fact, rhinos are my favourite animals and so last year I went to Africa with a colleague for a month to study them.</p> <p>The zoo is open every day and I work five different days each week. I live in a small flat twenty minutes away and I get up at ten to seven and start work at eight. The first thing I do when I get home at quarter past five is have a shower!</p>	<p><b>Task:</b></p> <ul style="list-style-type: none"> <li>Read the article about Ingrid McFarlane and then answer the</li> <li>For questions 1 – 7 mark A, B or C on your answer sheet.</li> </ul> <p><b>1</b> Ingrid would like to  <b>A</b> take some exams.  <b>B</b> earn more money.  <b>C</b> change her job.</p> <p><b>2</b> How does Ingrid feel about working in bad weather?  <b>A</b> She hates getting dirty.  <b>B</b> She doesn't mind it.  <b>C</b> She likes the snow.</p> <p><b>3</b> If Ingrid doesn't check the monkeys  <b>A</b> they may become ill.  <b>B</b> they may get hungry.  <b>C</b> they may run away.</p> <p><b>4</b> The animals Ingrid likes best are the  <b>A</b> monkeys.  <b>B</b> deer.  <b>C</b> rhinos.</p> <p><b>5</b> Ingrid travelled to Africa  <b>A</b> to have a month's holiday.  <b>B</b> to visit a colleague there.  <b>C</b> to learn more about some animals.</p> <p><b>6</b> The zoo is open  <b>A</b> only five days a week.  <b>B</b> seven days a week.  <b>C</b> on different days every week.</p> <p><b>7</b> Ingrid arrives at her flat in the evening at  <b>A</b> five fifteen.  <b>B</b> twenty past five.  <b>C</b> ten to seven.</p>
---	--

Abbildung 3: Aufgabenbeispiel 3 der Bildungsstandards für den Kompetenzbereich Leseverstehen im Fach Englisch in Anlehnung an KMK (2004d, S. 29 f.)

### Fiktives Beispiel: Die Bedeutung unterschiedlicher Eingangsvoraussetzungen der Schülerschaft

Seit langer Zeit ist bekannt, dass im Ruhrgebiet ein soziales Nord-Süd-Gefälle besteht. Der Norden des Ruhrgebietes beinhaltet tendenziell strukturschwächere, der Süden strukturstärkere Stadtteile. Diese Differenz der Sozialstruktur spiegelt sich auch im sozioökonomischen Hintergrund der Schülerschaft zweier fiktiver Schulen wider, der Nordschule und der Südschule (Tabelle 2, S. 11).

Betrachtet man die reinen numerischen Ergebnisse, so erzielt die Südschule deutlich bessere Leistungen als die Nordschule (Spalte 3). Um dem sozialen Hintergrund der Schülerschaft bei der Bewertung der schulischen Arbeit gerecht zu werden, wurden von allen Schülerinnen und Schülern zwei Sozialindikatoren erfragt: Die Anzahl von Einfamilienhäusern im Umkreis von 500 Metern um die eigene Wohnung und die Anzahl von Büchern im Haushalt der Familie (Spalte 7).

Es konnte statistisch erfasst werden, dass die Testleistung aller Schülerinnen und Schüler bei einer Vergleichsarbeit pro Einfamilienhaus im Wohnungsumfeld um (durchschnittlich) 0,5 Testpunkte und pro Buch im Haushalt um 0,02 Testpunkte zunimmt. Um einen fairen Vergleich zu ermöglichen, wurden die tatsächlichen Testergebnisse daher in zwei Schritten am jeweiligen Mittelwert der Anzahl von Einfamilienhäusern im Umkreis der eigenen Wohnung und der Anzahl von Büchern im Haushalt korrigiert. Im ersten Korrekturschritt wurde das Testergebnis für jedes zusätzliche Einfamilienhaus im Umkreis der eigenen Wohnung größer dem Mittelwert 4 um jeweils 0,5 Testpunkte reduziert und für jedes Einfamilienhaus kleiner 4 um die gleiche Punktzahl erhöht (Spalte 4). Es ist anhand der Mittelwerte pro Schule ersichtlich, dass sich die korrigierten Testleistungen der beiden Schulen bereits angenähert haben. Im zweiten Schritt wurde nun der Faktor „Bücher im Haushalt“ als Indikator für kulturelles Kapital berücksichtigt.

Dazu wurden die Testleistungen um die Differenz der sich im Haushalt befindlichen Bücher gegenüber dem Mittelwert von 90 Büchern korrigiert (Spalte 5). Nach diesem Korrekturschritt schneidet die Nordschule nun sogar besser ab als die Südschule. Berücksichtigt man also Unterschiede der Schülerschaft, so kann die Qualität der Beschulung in einem ganz anderen Licht erscheinen.

	Schüler	Testleistung	Korrektur 1	Korrektur 2	Anzahl der Einfamilienhäuser im Wohnumfeld	Anzahl der Bücher im Haushalt
<b>Nordschule</b>	Schüler 1	6	6	7,2	4	30
	Schülerin 2	4	5	6,5	2	15
	Schüler 3	10	12	12,8	0	50
	Schülerin 4	12	13	14,3	2	25
<b>Südschule</b>	Schüler 5	14	12	11,2	8	130
	Schülerin 6	16	15	13	6	190
	Schüler 7	10	9	7,2	6	180
	Schülerin 8	8	8	7,8	4	100
<b>Mittelwerte</b>	gesamt	10	10	10	4	90
	Nordschule	8	9	10,2	2	30
	Südschule	12	11	9,8	6	150

Tabelle 2: Fiktives Beispiel für die Adjustierung von Testleistungen aufgrund des sozioökonomischen Hintergrundes der Schülerschaft

Die statistische Korrektur von Leistungen für den sozialen Vergleich impliziert natürlich nicht, dass die Schülerinnen und Schüler der Nordschule in Bezug auf das Lehrziel „mehr können“, als sie im Test gezeigt haben. Die durchschnittliche Anzahl gelöster Items liegt in der Nordschule bei 8, und die Schülerinnen und Schüler der Nordschule haben tatsächlich im Sinne eines kriterialen Vergleichs „Aufholbedarf“ gegenüber der Südschule. Dieser „Aufholbedarf“ ist aber nicht das Resultat ungünstiger Beschulung, sondern ungünstiger Eingangsvoraussetzungen der Schülerschaft an der Nordschule.

Um „faire Vergleiche“ unter Berücksichtigung leistungsrelevanter Kontextmerkmale der Schülerschaft zu realisieren, lassen sich zwei grundlegende Konzepte<sup>4</sup> unterscheiden (vgl. für eine ausführlichere Darstellung Nachtigall, Kröhne, Enders & Steyer, 2008): faire Vergleiche – Methoden

1. Der Vergleich mit Klassen (bzw. Schulen), deren Schülerschaft möglichst große Ähnlichkeit in Bezug auf sozioökonomische Merkmale besitzt. Schulen werden dazu sogenannten Belastungs-, Kontext- oder Standortgruppen zugeordnet.
2. Der Vergleich mit statistisch korrigierten Leistungswerten, die tatsächliche Schulleistungen in Beziehung zu Schulleistungen setzen, die aufgrund des sozioökonomischen Hintergrundes der Schülerschaft zu erwarten wären. Dazu werden die Testergebnisse statistisch „aufgewertet“, wenn die Schülerzusammensetzung der Klasse ungünstiger als in Referenzklassen ist. Die Testergebnisse werden hingegen „abgewertet“, wenn die Zusammensetzung der Schülerschaft gegenüber Referenzklassen als günstiger zu bezeichnen ist.

Das erste vorgestellte Konzept bietet Lehrerinnen und Lehrern insbesondere den Vorteil höherer Transparenz und Nachvollziehbarkeit des sozialen Vergleichs, beinhaltet aber auch die Herausforderung, geeignete Vergleichsklassen oder -schulen zu identifizieren. Bei Anwendung des zweiten Konzepts lassen sich genauere Ergebnisse ermitteln; deren Zustandekommen ist für die Adressaten der Rückmeldung aber häufig wenig durchschaubar. Zu beiden Konzepten ein kurzes Beispiel:

### Beispiele:

Im Rahmen der Ergebnissrückmeldung der Lernstandserhebungen in der achten Jahrgangsstufe in Nordrhein-Westfalen (lernstand 8) erhalten Schulen als Referenzwerte für den sozialen Vergleich die Testergebnisse der jeweiligen Schulform sowie die Testergebnisse so genannter Standorttypen. Jede Schule ist auf Basis von Daten der amtlichen Statistik<sup>5</sup> einem von insgesamt fünf Standorttypen zugeordnet worden. Als relevante Daten wurden berücksichtigt: faire Vergleiche – Beispiele

- Anteil der Migrantinnen und Migranten innerhalb der Schülerschaft

<sup>4</sup> Nachtigall, Kröhne, Enders und Steyer (2008) unterscheiden neben dem Vergleich mit dem (ungewichteten) Populationsmittelwert drei Verfahren: ‚comparison to similar existing classes or schools‘, ‚comparison to expected values‘ und ‚comparison based on propensity scores‘. Aus didaktischen Gründen sind an dieser Stelle die letzten beiden Verfahren zusammengefasst worden.

<sup>5</sup> Den zugrundeliegenden Kernel-Density-Ansatz beschreibt Schröpler (2009).

- Anteil der Arbeitslosen und der SGB II-Empfängerinnen und Empfänger unter 18 Jahren im Schulumfeld

Schulen des Standorttyps 1 sind durch einen niedrigen Anteil von Empfängerinnen und Empfängern staatlicher Sozialhilfeeleistungen, von Arbeitslosen und von Menschen mit Migrationshintergrund charakterisiert; Schulen des Standorttyps 5 befinden sich hingegen eher in einer Lage mit schwierigen sozialen und kulturellen Rahmenbedingungen (Isaac, 2011). Sozialräumliche Veränderungen, wie etwa in der Bevölkerungsstruktur, werden durch regelmäßige Aktualisierung der amtlichen Daten berücksichtigt.

Bei den Kompetenztests (<http://www.kompetenztest.de>) in den sechsten und achten Klassen in Thüringen werden die durchschnittlichen Testergebnisse von Klassen mit adjustierten Landesmittelwerten verglichen. Bei den Schülerinnen und Schülern werden die folgenden Merkmale zum Zweck der Adjustierung der individuellen Testleistung erhoben (Nachtigall, Kröhne & Müller, 2005):

- Geschlecht
- Muttersprache
- Wiederholung einer Klassenstufe
- Besondere Lernschwierigkeiten bzw. sonderpädagogischer Förderbedarf
- Anzahl der Bücher im Elternhaus

Auf Basis dieser Werte wird für jede Schülerin und jeden Schüler ein Erwartungswert<sup>6</sup> berechnet; der durchschnittliche Erwartungswert in einer Klasse wird als Indikator für die zu erwartende Klassenleistung („korrigierter Landesmittelwert“) herangezogen. Lehrerinnen und Lehrer in Thüringen erhalten dementsprechend einen individuellen, auf die besondere Zusammensetzung ihrer Klasse abgestimmten Erwartungswert der Leistungsstärke ihrer Lerngruppe.

### 2.2.3 Individuelle Bezugsnorm bei Vergleichsarbeiten

individuelle  
Bezugsnorm;  
regelmäßige  
Lernstands-  
diagnose

Auf der Ebene der Einzelschule lassen sich die Ergebnisse der Vergleichsarbeiten für die Evaluation schulischer Entwicklungsprozesse nutzen, wenn zusätzliche diagnostische Informationen berücksichtigt werden. Materialien zur regelmäßigen Lernstandsdiagnose können an von Lehrkräften selbst gewählten Zeitpunkten für die Überprüfung der Leistungsentwicklungen und Kompetenzzuwächse eingesetzt werden (vgl. hierzu Kapitel 2.5).

### 2.2.4 Weiterführende Literatur

Literatur

Die Bildungsstandards als zentrale kriteriale Bezugsnorm von Vergleichsarbeiten können über die Webseiten der KMK als Online-Dokumente bezogen werden (siehe Kapitel 2.4). Das Lehrbuch von Klauer (1987) gilt als Klassiker für die Darstellung kriteriumsorientierten Testens in der Schule. Faire Schulvergleiche thematisiert Arnold (1999). Die Bedeutung von Hintergrundmerkmalen für die Entstehung von Leistungsheterogenität und Möglichkeiten fairer Vergleiche werden von Nachtigall, Kröhne, Enders & Steyer (2008) dargestellt.

### 2.2.5 Verständnis und Diskussionspunkte

1. *Wie ist die Passung zwischen den Aufgaben der Vergleichsarbeiten und den Lehrplänen zu bewerten? Inwiefern könnte eine grobe, indirekte Anbindung vorliegen?*
2. *Bei Parallelarbeiten werden Aufgaben einer schriftlichen Leistungsüberprüfung von den Fachlehrerinnen und Fachlehrern einer Jahrgangsstufe gemeinsam entwickelt und in Parallelklassen eingesetzt. Die Auswertung der Aufgabenlösungen erfolgt wiederum in Absprache zwischen den Fachlehrerinnen und Fachlehrern. Diskutieren Sie bitte die Bedeutung von Parallelarbeiten vor dem Hintergrund der sozialen Bezugsnorm. Wie können sich Vergleichsarbeiten und Parallelarbeiten gegenseitig ergänzen?*
3. *Denken Sie sich bitte in folgendes Szenario: Um eine kostenintensive Befragung von Schülerinnen und Schülern zu vermeiden, sollen die relativ leicht zugänglichen Daten der Schulstatistik herangezogen werden, um eine Adjustierung von Schulleistungsdaten (z.B. aus Vergleichsarbeiten) vornehmen zu können und den Schulen damit faire Vergleiche zu ermöglichen. Welchen Kriterien sollten die Daten der Schulstatistik genügen? Welche Schwierigkeiten können entstehen, wenn solche Daten verwendet werden?*

<sup>6</sup> Die Erwartungswerte werden über die Zellenmittelwerte einer mehrfaktoriellen Varianzanalyse geschätzt, siehe Nachtigall, Kröhne, Enders & Steyer (2008).

## 2.3 Testkonstruktion

Welche Bedeutung haben Kompetenzstufen und was sagen sie aus? Wie werden Schülerinnen und Schüler bei der Ergebnissrückmeldung von Vergleichsarbeiten diesen Kompetenzstufen zugeordnet? Woher weiß man, dass die Schülerinnen und Schüler die jeweiligen Aufgaben dieser Stufen mit einiger Sicherheit lösen können? Im folgenden Kapitel wird erläutert, wie aufeinander aufbauende Leistungsanforderungen durch Kompetenzstufen beschrieben und mit Hilfe sorgfältig ausgewählter Aufgaben erfassbar gemacht werden. Bei der Aufgabenerstellung und -auswahl steht die Passung in Bezug auf die Anforderungen der Bildungsstandards im Vordergrund. Außerdem wird erläutert, wie man auf Basis des Testmodells nach Rasch (Rasch, 1960) sicherstellt, dass Schülerinnen und Schüler den Anforderungen ihrer Kompetenzstufe mit einiger Sicherheit gewachsen sind. Die aus Studienbrief 1 bekannten Konzepte der Testgüte werden an verschiedenen Stellen aufgegriffen.

In diesem Kapitel werden folgende Fragen beantwortet:

- *Wie werden Leistungsanforderungen durch Kompetenzstufen beschrieben?*
- *Wie werden Schülerinnen und Schüler den Kompetenzstufen zugeordnet? Welchen Beitrag leistet das Rasch-Modell zur Konstruktion von Kompetenzskalen?*
- *Nach welchen Kriterien werden die Testaufgaben für Vergleichsarbeiten ausgewählt und wie wird sichergestellt, dass diese Kriterien eingehalten werden?*

### 2.3.1 Die Beschreibung von Leistungsanforderungen durch aufeinander aufbauende Kompetenzstufen

Die in den Bildungsstandards festgeschriebenen Kompetenzerwartungen werden durch eigens für diesen Zweck entwickelte und bzgl. ihrer Testgüte (vgl. Kapitel 1.3) sorgfältig überprüfte Aufgaben erfasst. Bei der Ergebnissrückmeldung werden Aufgaben, die ähnlich komplexe Anforderungen beinhalten, entsprechend ihrer Schwierigkeit gruppiert und als Kompetenzstufen inhaltlich beschrieben. Die Kompetenzstufen bauen in ihren Anforderungen aufeinander auf, d.h. Schülerinnen und Schüler, die höhere Kompetenzstufen erreicht haben, beherrschen auch die Inhalte niedrigerer Stufen sicher und lösen die Aufgaben entsprechender Kompetenzanforderungen mit hoher Wahrscheinlichkeit. Beispielhaft sei dies im Folgenden an der Beschreibung eines Kompetenzstufenmodells aus VERA (Vergleichsarbeiten in der dritten Jahrgangsstufe), Kompetenzbereich „Sachrechnen und Größen“, verdeutlicht (Helmke & Hosenfeld, 2004). Für diesen Kompetenzbereich der Primarstufen-Mathematik waren zu diesem Zeitpunkt noch drei Stufen voneinander abgegrenzt worden, die eine zunehmende Komplexität der Anforderungen (Progression) erkennen lassen.

Prüfung der  
Testgüte

#### Beispiel: Erläuterungen zu den Kompetenzstufen Mathematik: „Sachrechnen und Größen“ bei VERA 2004

##### Stufe 1: Elementare Kenntnisse

- Die Anwendung von Addition (auch wiederholte Additionen) und Subtraktion in authentischen Aufgaben gelingt bei Aufgaben mit Auswahl aus vorgegebenen Lösungen.
- Grundlegende Kenntnisse von vertrauten Maßeinheiten (Längen-, Zeit-, Gewichts- und Geldeinheiten).
- Offensichtlich unlösbare Aufgaben werden erkannt.

##### Stufe 2: Entwickelte Fähigkeiten im Umgang und Rechnen mit Größen

- Im Umgang mit vertrauten Maßeinheiten (Längen-, Zeit-, Gewichts- und Geldeinheiten) können Aufgaben bis in den Tausender-Zahlenraum gelöst werden.
- Lösungen von authentischen Aufgaben, die Umrechnungen von Maßeinheiten erfordern, gelingen.
- Rundungen und Schätzungen gelingen bei Aufgaben mit vorgegebenen Lösungen.
- Verknüpfungen von Operationen werden bewältigt.
- Der Umgang mit elementaren Brüchen gelingt.
- Aufgaben mit mehreren zu verarbeitenden Größen werden gemeistert.

##### Stufe 3: Eigenständige Problemlösungen

- Unlösbare Aufgaben, die eine mentale Vorstellung des geschilderten Szenarios erfordern, werden erkannt.
- Bei Aufgaben ohne vorgegebene Fragestellung kann eigenständig eine Aufgabe formuliert und bearbeitet werden.
- Funktionale Beziehungen zwischen Maßen können eigenständig hergestellt und verglichen werden.
- Die mathematische Modellierung problemhaltiger Sachsituationen gelingt.
- Aufgaben, die mehrere Teilschritte umfassen, werden beherrscht.

Während Schülerinnen und Schüler auf Stufe 1 beispielsweise lediglich „Strichrechnung“ sicher anwenden können und grundlegende Kenntnisse über Größen und Maßeinheiten besitzen, können auf Stufe 2 bereits Brüche bearbeitet werden, werden Maßeinheiten in einem breiten Zahlenraum beherrscht und mehrere Operationen und Größen verknüpft. Schülerinnen und Schüler auf Stufe 3 sind darüber hinaus in der Lage, Größen und Maße zueinander in Beziehung zu setzen, eigenständige Modellierungen vorzunehmen und mehrschrittige Aufgaben zu lösen.

### 2.3.2 Definition von Kompetenzstufen über Lösungswahrscheinlichkeiten: Das Rasch-Modell

**Rasch-Modell** Für die Definition von Kompetenzstufen ist es wichtig sicherzustellen, dass Schülerinnen und Schüler der richtigen Kompetenzstufe zugeordnet werden, also die jeweiligen Anforderungen bewältigen und entsprechende Aufgaben hinreichend sicher lösen. Im sogenannten Rasch-Modell (Rasch, 1960) aus der probabilistischen Testtheorie (Rost, 2004) wird die Wahrscheinlichkeit einer korrekt gelösten Aufgabe als eine (logistische) Funktion der Schülerkompetenz ( $\theta$ ) und der Schwierigkeit der Aufgabe ( $\sigma$ ) geschätzt und kann mit Hilfe einer so genannten Itemcharakteristischen Kurve (ICC) veranschaulicht werden. Beispielhaft werden in Abbildung 4 die ICCs für zwei Testaufgaben dargestellt. In der Abbildung sind die Aufgabenschwierigkeit bzw. Schülerkompetenz (beide werden im Rasch-Modell auf einer Skala abgebildet) gegen die Lösungswahrscheinlichkeit aufgetragen. Es ist ersichtlich, dass die ICCs für beide Aufgaben dieselbe Steigung besitzen. Aufgabe 2 ist also - unabhängig von der Schülerkompetenz, die hier beispielhaft für einen Schüler 1 mit geringer und einen Schüler 2 mit hoher Kompetenzausprägung veranschaulicht wurde - stets die schwierigere der beiden Testaufgaben. Als Aufgabenschwierigkeit im Rasch-Modell wird bei Vergleichsarbeiten zumeist derjenige Punkt der ICC gewählt, bei dem die Lösungswahrscheinlichkeit für die Aufgabe bei 62,5 % liegt. Um die Anforderungen der Aufgabe 1 sicher zu lösen, müsste ein Schüler auf der Kompetenzskala also (mindestens)  $\theta = -2,2$  erreichen, für die schwierigere Aufgabe 2 mindestens  $\theta = 0,7$ . Eine bemerkenswerte Eigenschaft des Rasch-Modells ist die Tatsache, dass für die Schätzung der Kompetenzausprägung der Schülerinnen und Schüler  $\theta$  lediglich die Summe gelöster Items unabhängig von deren Schwierigkeit relevant ist.

Die Kompetenzstufen werden - einem Konzept aus dem PISA-Kontext folgend - so gebildet, dass Schülerinnen und Schüler mit der niedrigsten Kompetenz auf jeder Stufe (niedrigste Schülerkompetenz innerhalb dieser Stufe) eine Lösungswahrscheinlichkeit von mindestens 50 % für alle Aufgaben der Stufe und eine Lösungswahrscheinlichkeit von 62,5 % für die leichteste Aufgabe der Stufe besitzen (in Abbildung 5 in blauer Farbe dargestellt). Durchschnittlich kompetente Schülerinnen und Schüler dieser Stufe (Mittelwert der Schülerkompetenz innerhalb der Kompetenzstufe) lösen eine durchschnittlich schwierige Aufgabe der gleichen Stufe mit einer Wahrscheinlichkeit von 62,5 % (in Abbildung 5 in roter Farbe dargestellt). Es wird so sichergestellt, dass Aufgaben einer Kompetenzstufe von den Schülerinnen und Schülern dieser Stufe mit einiger Sicherheit gemeistert werden.

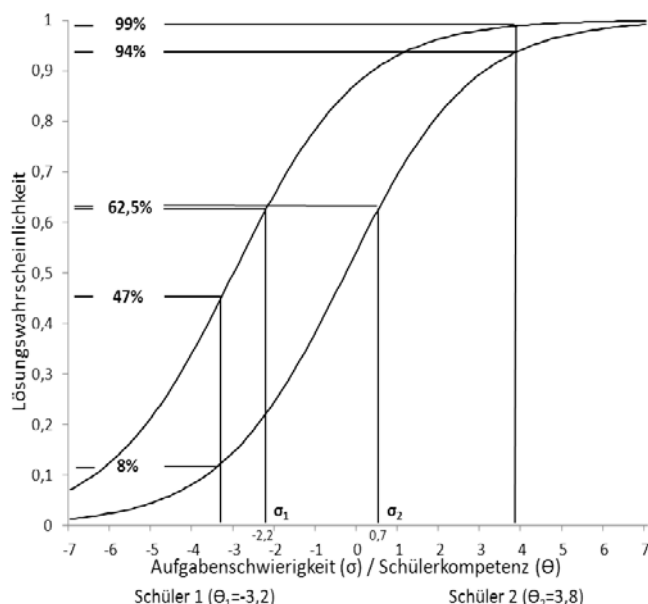


Abbildung 4: Itemcharakteristische Kurven (ICC) zweier Testaufgaben nach dem dichotomen Rasch-Modell – Zusammenhang von Aufgabenschwierigkeit ( $\sigma$ ), Schülerkompetenz ( $\theta$ ) und Lösungswahrscheinlichkeit

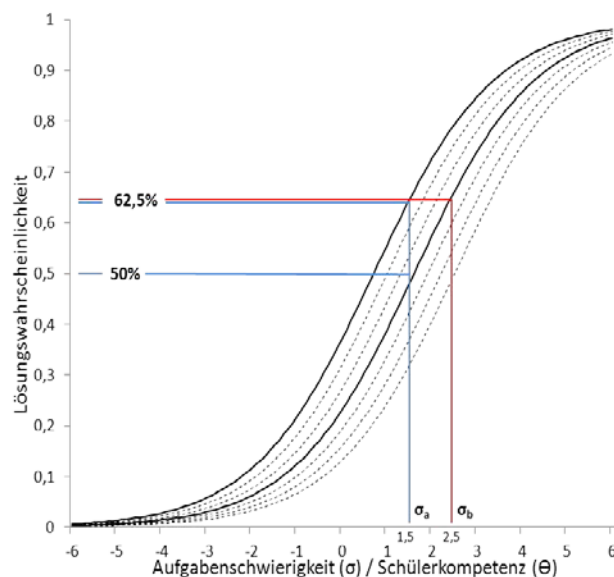


Abbildung 5: Lösungswahrscheinlichkeiten für Aufgaben einer Kompetenzstufe durch Schülerinnen und Schüler mit der niedrigsten Kompetenz innerhalb der Stufe (in blauer Farbe) und Schülerinnen und Schüler durchschnittlicher Kompetenz innerhalb der Stufe (in roter Farbe). Die Aufgabe mit der niedrigsten Schwierigkeit und die Aufgabe mittlerer Schwierigkeit sind durchgehend gezeichnet.



### 2.3.3 Aufgabenauswahl bei Vergleichsarbeiten: Orientierung an Bildungsstandards und hohe Testgüte

Die Qualität eines Kompetenzstufenmodells ist auch von der Auswahl geeigneter Testaufgaben abhängig. Die Aufgaben der Vergleichsarbeiten sollen die Bildungsstandards widerspiegeln, dürfen die jeweilige fachdidaktische Tradition nicht konterkarieren und müssen über eine entsprechende Testgüte verfügen (vgl. Burkard & Peek, 2004; Dobbelstein & Peek, 2007; Fleischer, Wirth & Leutner, 2007). Hinzu kommt, dass die intendierten sozialen Vergleiche (Kapitel 2.2) nur sinnvoll sind, wenn Lehrkräfte gleiche Aufgabenlösungen gleich auswerten, die Aufgaben also objektiv (auswertbar) sind. Um diesen Ansprüchen gerecht werden zu können, werden alle Testaufgaben unter realen Bedingungen mehrschrittig erprobt: Der erste Schritt beinhaltet eine Präpilotierung, bei der die Schwierigkeit der Aufgaben und die Güte der Auswertungsanleitungen grob abgeschätzt werden. Darauf folgt eine Pilotierungsphase, in der eine ausreichend große Stichprobe von Schülerinnen und Schülern die Testaufgaben bearbeitet. Die Testgüte der Aufgaben wird folgendermaßen sichergestellt:

1. Die Objektivität der Aufgabenbewertungen wird durch eine hohe Anzahl von Aufgaben geschlossener Formate erhöht. Zudem kann mit Hilfe statistischer Indizes (Wirtz & Caspar, 2002) die Übereinstimmung von mindestens zwei Lehrkräften bei der Beurteilung von Aufgabenlösungen derselben Person erfasst werden.
2. Wie auch in der klassischen Testtheorie (siehe Studienbrief 1) wird die Reliabilität der aus den Aufgaben zusammengesetzten Skala geprüft. Tests der bei Vergleichsarbeiten überwiegend eingesetzten probabilistischen Testtheorie (siehe oben) nutzen dazu Methoden analog zum in Studienbrief 1 vorgestellten Koeffizienten Cronbachs Alpha. Die Reliabilität der Vergleichsarbeiten ist hinreichend hoch, um zuverlässige diagnostische Aussagen auf der Ebene von Schulklassen zu treffen. Individualdiagnostische Aussagen (vgl. Studienbrief 1) verbieten sich allerdings weitestgehend (Leutner, Fleischer, Spoden & Wirth, 2007). Um Kompetenzen einzelner Schülerinnen und Schüler für diese Zielsetzung genau genug zu erfassen, wäre eine höhere Item-Anzahl notwendig. Dies sollte bei der Ergebnismrückmeldung an die Schülerinnen und Schüler beachtet werden.
3. Die Validität kann – wenn überhaupt – nur grob erfasst werden. Vereinzelt kann aber geprüft werden, ob sich in den Ergebnissen bekannte Befunde (z.B. aus dem Bildungsmonitoring, vgl. Studienbrief 3) widerspiegeln, etwa Schulform- oder Geschlechtsunterschiede sowie Zusammenhänge mit sozioökonomischen Variablen.

Die Aufgabensätze der Vergleichsarbeiten stellen schließlich einen Kompromiss zwischen fachdidaktischen und testtheoretischen Anforderungen dar, d.h. inhaltliche Qualität und Testgüte müssen gleichermaßen gewährleistet sein.

Das deutsche Schulsystem weist eine breite Leistungsstreuung zwischen Schulen (insbesondere zwischen Schulen unterschiedlicher Schulformen), aber eine verhältnismäßig geringe Streuung innerhalb der Einzelschule auf (Baumert, Trautwein & Artelt, 2003). Diesem Ergebnis wird bei Vergleichsarbeiten einerseits durch die Verwendung unterschiedlich schwieriger Testheftversionen, andererseits durch ein breites Schwierigkeitsspektrum der Testaufgaben innerhalb der Hefte Rechnung getragen. Es wird so sichergestellt, dass die Testaufgaben Schülergruppen nicht systematisch über- oder unterfordern. Innerhalb einer Lerngruppe werden identische Aufgaben verwendet, sodass gut nachvollziehbare vergleichende Aussagen über die getesteten Kompetenzen getroffen werden können. Ein typisches Missverständnis bei der Interpretation von Ergebnissen aus Vergleichsarbeiten verbirgt sich in der Annahme, dass Leistungen über verschiedene Schulformen hinweg nicht vergleichbar seien. Dies trifft auf Vergleichsarbeiten in der Regel nicht zu<sup>7</sup>. Bei der Testauswertung auf Basis des zuvor dargestellten Rasch-Modells sind die Ergebnisse von Klassen unterschiedlicher Schulformen auch dann noch vergleichbar, wenn in den Schulformen unterschiedliche Testhefte mit zum Teil anderen Aufgaben bearbeitet wurden.

### 2.3.4 Weiterführende Literatur

Das Vorgehen bei der Aufgabenentwicklung und Aufgabenselektion wird ausführlich von Leutner, Fleischer, Spoden & Wirth (2007) beschrieben. Das Rasch-Modell erläutern Bond & Fox (2001) weitestgehend „formelfrei“ und Rost (2004) auf Schulmathematik-Niveau. Die Definition von Kompetenzstufen auf der Rasch-Skala wird von Fleischer, Wirth & Leutner (2007) dargelegt, kann darüber hinaus aber auch den PISA-Bänden (z.B. Artelt, Stanat, Schneider & Schiefele, 2001) oder ebenfalls Leutner, Fleischer, Spoden & Wirth (2007) entnommen werden.

### 2.3.5 Verständnis und Diskussionspunkte

1. *Fassen Sie bitte zusammen, nach welchen Kriterien die Qualität der Aufgaben in Vergleichsarbeiten geprüft wird.*
2. *Beschreiben Sie bitte, weshalb Testaufgaben bei Vergleichsarbeiten objektiv auswertbar sein sollen. Überlegen Sie, wie Sie selbst – ohne Aneignung weiteren Wissens – die Auswertungsobjektivität von Aufgaben (z.B. bei Parallelarbeiten) überprüfen könnten.*
3. *Diskutieren Sie bitte, welche Informationen als Hinweise auf die Validität eines für Vergleichsarbeiten konstruierten Tests herangezogen werden könnten.*

<sup>7</sup> Voraussetzung ist allerdings, dass ein gemeinsames Kompetenzstufenmodell zugrunde liegt. Derzeit stellt das Kompetenzstufenmodell für den mittleren Schulabschluss (MSA) den Referenzrahmen dar.

## 2.4 Inhaltlicher Anwendungsbereich/Phänomenbereich<sup>8</sup>

VERA 3/  
VERA 8

Vergleichsarbeiten werden in der dritten Jahrgangsstufe (VERA 3) in den Fächern Deutsch und Mathematik geschrieben. In der achten Jahrgangsstufe finden Vergleichsarbeiten (VERA 8) in den Fächern Deutsch, Mathematik und der ersten Fremdsprache (Englisch oder Französisch) statt. Die Webseite des IQB ([http://www.iqb.hu-berlin.de/vera?lang=en&reg=r\\_1](http://www.iqb.hu-berlin.de/vera?lang=en&reg=r_1)) weist Vergleichsarbeiten in allen Bundesländern der BRD aus. Aktuelle Informationen zu Vergleichsarbeiten in jedem Bundesland sind den beim IQB aufgeführten Webseiten der Länder zu entnehmen.

Dieses Kapitel umfasst Hintergrundinformationen zu den Kompetenzstufenmodellen in zwei bzw. drei Fächern der Primar- und Sekundarstufe<sup>9</sup>. Die Leserinnen und Leser sind eingeladen, die jeweils für sie relevanten Unterkapitel auszuwählen:

2.4.1.1: Vergleichsarbeiten in der dritten Jahrgangsstufe (VERA 3) im Fach Deutsch

2.4.1.2: Vergleichsarbeiten in der dritten Jahrgangsstufe (VERA 3) im Fach Mathematik

2.4.2.1: Vergleichsarbeiten in der achten Jahrgangsstufe (VERA 8) im Fach Deutsch

2.4.2.2: Vergleichsarbeiten in der achten Jahrgangsstufe (VERA 8) im Fach Englisch

2.4.2.3: Vergleichsarbeiten in der achten Jahrgangsstufe (VERA 8) im Fach Mathematik

In diesem Kapitel werden folgende Fragen beantwortet:

- Welche Inhalte werden bei den Vergleichsarbeiten in der dritten Jahrgangsstufe (VERA 3) in den Fächern Deutsch und Mathematik thematisiert?
- Wie werden diese Inhalte bei VERA 3 operationalisiert und welche Aufgabenformate kommen zum Einsatz?
- Welche Inhalte werden bei den Vergleichsarbeiten in der achten Jahrgangsstufe (VERA 8) in den Fächern Deutsch, Mathematik und der ersten Fremdsprache thematisiert?
- Wie werden diese Inhalte bei VERA 8 operationalisiert und welche Aufgabenformate kommen zum Einsatz?

### 2.4.1 Vergleichsarbeiten in der dritten Jahrgangsstufe (VERA 3)

VERA 3

Vergleichsarbeiten sind seit 2004 in nahezu allen Bundesländern in der Primarstufe installiert worden. Seit 2007 werden Vergleichsarbeiten in der Primarstufe einheitlich gegen Ende der dritten Jahrgangsstufe geschrieben, sodass die Schulen eine Rückmeldung darüber erhalten, wie weit ihre Schülerinnen und Schüler von den durch die KMK (KMK, 2004a-b) in den Bildungsstandards veröffentlichten Kompetenzerwartungen am Ende der Jahrgangsstufe vier entfernt sind. Neben dem Angebot des IQB bieten die Webseiten der Projektgruppe VERA an der Universität Koblenz-Landau, welche derzeit für acht Bundesländer ein gemeinsames Rückmeldesystem koordiniert, Beispielaufgaben und zahlreiche Materialien aus den Durchführungen 2004-2009 zum Download an (<http://www.uni-landau.de/vera/>).

#### 2.4.1.1 Kompetenzskalen im Fach Deutsch bei VERA 3

Kompetenz-  
bereiche  
im Fach  
Deutsch

„Sprache ist Träger von Sinn und Überlieferung, Schlüssel zum Welt- und Selbstverständnis und Mittel zwischenmenschlicher Verständigung. Sie hat grundlegende Bedeutung für die kognitive, emotionale und soziale Entwicklung der Kinder“, heißt es in den Bildungsstandards für die Primarstufe im Fach Deutsch (KMK, 2004a).

Für die Entwicklung von Kompetenzstufenmodellen für VERA wurden theoretische und empirische Erkenntnisse, beispielsweise aus den IGLU-Erhebungen (Internationale Grundschul-Lese-Untersuchung; Bos et al., 2003, 2007), LAU (Aspekte der Lernausgangslage und Lernentwicklung; Lehmann & Peek, 1997) und DESI (Deutsch Englisch Schülerleistungen International; DESI-Konsortium, 2008) aufgearbeitet. Die resultierenden Aufgabenmengen für das Fach Deutsch operationalisiert aus den Bildungsstandards die Kompetenzbereiche

- Leseverstehen,
- Hörverstehen,
- Sprache und Sprachgebrauch untersuchen,
- Schreiben und
- Orthografie.

Die Durchführung der VERA-Tests umfasst zwei Testtage, wobei der Kompetenzbereich Leseverstehen jeweils am ersten Testtag, einer der weiteren Kompetenzbereiche am zweiten Testtag erfasst wird.

<sup>8</sup> Stand: 22.06.2011; vgl. [http://www.iqb.hu-berlin.de/bista?reg=r\\_4](http://www.iqb.hu-berlin.de/bista?reg=r_4).

<sup>9</sup> Die nachfolgende Darstellung der Kompetenzstufenmodelle bei VERA 3 und VERA 8 spart notwendigerweise Details aus. Die Beschreibung der Vergleichsarbeiten in der ersten Fremdsprache wurde auf das Fach Englisch beschränkt, in den Fächern Deutsch und Englisch wurde jeweils ein Kompetenzbereich ins Zentrum gerückt.

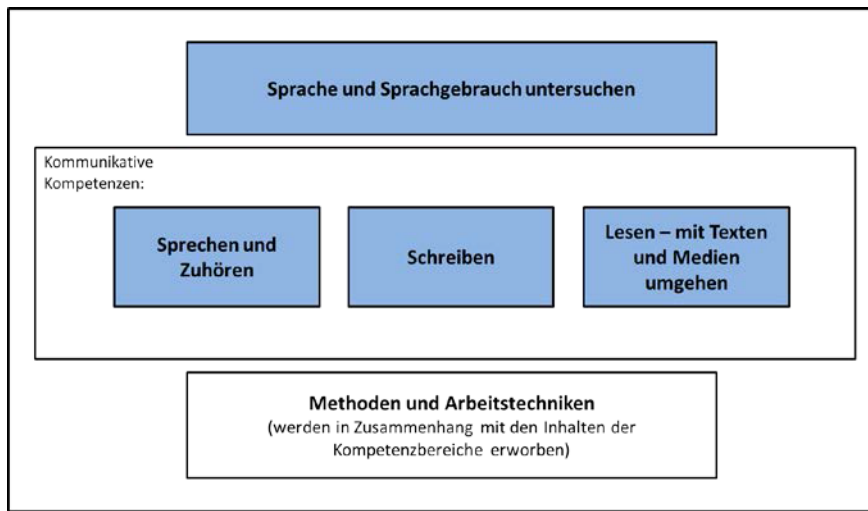


Abbildung 6: Kompetenzbereiche der Bildungsstandards für die Primarstufe im Fach Deutsch

Hinweis: Blau unterlegt jene Kompetenzbereiche, für die Kompetenzstufenmodelle (teilweise in Bezug auf Subkomponenten) entwickelt wurden.

Rechtschreibung spielt außerhalb der Kompetenzbereiche „Schreiben“ und „Orthografie“ eine untergeordnete Rolle. Entsprechend kommen in den anderen Kompetenzbereichen zumeist geschlossene Aufgabenformate wie Zuordnungs- oder Multiple-Choice-Aufgaben zum Einsatz, die geringe Textproduktion beinhalten. In dieser Entscheidung spiegelt sich das Bestreben wider, Teilleistungsbereiche der Fächer valide abzubilden. Die differenzierte und objektive Erfassung von Schreibleistungen im Rahmen von Vergleichsarbeiten stellt nach wie vor eine große Herausforderung dar. Die Schülerinnen und Schüler erhalten hier eine offene, aber klar umschriebene Aufgabenstellung zur Formulierung eines Textes. Die Auswertung findet anhand von Indikatoren statt, welche die Erfüllung von text- bzw. aufgabenspezifischen und allgemeinen Kompetenzmerkmalen (Anzahl der Wörter, Anzahl richtig geschriebener Wörter, Morphematik, Syntax und Textverknüpfungen) erfragen und somit einen indirekten Schluss auf Schreibkompetenzen der Schülerinnen und Schüler zulassen. Tabelle 3 zeigt ein Beispiel der VERA 3-Erhebung 2009 für einen solchen Indikator des Kompetenzbereichs „Schreiben“ (Projekt VERA, 2009).

Inhalte der Kompetenzbereiche und Aufgabenformate

Kriterium	Punktevergabe	Beschreibung	Beispiel
nicht erfüllt	0	<ul style="list-style-type: none"> <li>Es gibt keine konsistente Perspektive.</li> <li>Im Text werden mindestens 3 verschiedene Perspektiven eingenommen, z.B. Ich-Wir, Sie-Du, Ihr-Die Eltern.</li> <li>Ob ein Wechsel zur neutralen Perspektive (man) einen Bruch darstellt, hängt von der Perspektive des restlichen Testes ab.</li> </ul>	„Es wäre nett, dass <b>SIE</b> und die Kinder kommen. Natürlich könnt <b>IHR</b> auch Bekannte einladen [...] Essen und Trinken sollte <b>MAN</b> mitbringen...“
teilweise erfüllt	1	<ul style="list-style-type: none"> <li>Die Adressaten werden nicht durchgängig angesprochen</li> <li>Es werden zwei verschiedene Perspektiven eingenommen, gleichgültig wie oft zwischen ihnen gewechselt wird</li> <li>Ob ein Wechsel zur neutralen Perspektive (man) einen Bruch darstellt, hängt von der Perspektive des restlichen Textes ab.</li> </ul>	„Bei unserem Programm pusten <b>WIR</b> Luftballons auf, dann essen <b>WIR</b> , dann basteln <b>WIR</b> Piratensachen, dann hört <b>MAN</b> eine Geschichte, dann singen <b>WIR</b> .“
voll erfüllt	2	<ul style="list-style-type: none"> <li>Die Adressaten werden durchgängig angesprochen.</li> <li>Die Schreibperspektive (Schüler, Klasse, Lehrkraft) wird konsequent beibehalten. Dabei kann es sich auch um eine neutrale Perspektive handeln, in der niemand direkt angesprochen wird. Ob ein Wechsel zur neutralen Perspektive (man) einen Bruch darstellt, hängt von der Perspektive des restlichen Textes ab.</li> </ul>	„ <b>WIR</b> haben ein tolles Programm vorbereitet mit vielen Spielen. <b>MAN</b> kann dabei auch Preise gewinnen und Spaß haben.“

Tabelle 3: Beispiel für einen Indikator des Kompetenzbereichs „Schreiben“ im Fach Deutsch aus der VERA 3-Erhebung 2009 (Projekt VERA, 2009)



Im Folgenden wird beispielhaft der Kompetenzbereich „Sprache und Sprachgebrauch untersuchen“ beschrieben.

### Beispiel:

Beispiel  
Kompetenz-  
bereich  
„Sprache  
und Sprach-  
gebrauch  
untersuchen“

„Sprache und Sprachgebrauch untersuchen“ findet sich als Kompetenzbereich sowohl in den Bildungsstandards der Primarstufe als auch in den Standards für den Hauptschul- und mittleren Bildungsabschluss der Sekundarstufe I. Er überspannt die kommunikativen Kompetenzen Lesen, Hören, Schreiben und Sprechen, für die grundlegende Kompetenzen im Bereich „Sprache und Sprachgebrauch untersuchen“ eine Voraussetzung darstellen. Der Inhalt der Kompetenzen in der Primarstufe bezieht sich auf den bewussten Umgang mit und die bewusste Anwendung von Sprache sowie auf ein grundlegendes Wissen über grammatische Strukturen und die Bedeutung von Wörtern, Sätzen und Texten. Aufgaben des Kompetenzbereichs sollen Schülerinnen und Schülern anregen, Sprache in Bezug auf ihre inhaltliche Dimensionen und Verwendungszusammenhänge zu untersuchen. Dabei kann die Grundschule an frühe Spracherfahrungen anknüpfen, denn schon vor der Einschulung haben die Kinder eigenaktiv Wissen über Sprache erworben, und auch außerhalb des Unterrichts wird dieses Wissen parallel zur Schule weiterentwickelt. Es ist dabei zielführend, zwischen deklarativem (Wissen über Sprache) und prozeduralem Wissen (Fertigkeiten im Umgang mit Sprache) zu unterscheiden. Beide Wissensbereiche müssen nicht übereinstimmen, sondern können bei den Kindern sehr unterschiedlich stark ausgeprägt sein. Tabelle 4 zeigt beispielhaft die Inhalte der Teilkompetenzen „An Wörtern, Sätzen, Texten arbeiten“ sowie „Grundlegende sprachliche Strukturen und Begriffe kennen und verwenden“, die bei VERA ins Zentrum des Kompetenzbereichs gerückt worden sind - nicht zuletzt, da sie sich derzeit am besten operationalisieren lassen.

Teilkompetenz		Inhalte
<b>an Wörtern, Sätzen, Texten arbeiten, z.B.</b>		<ul style="list-style-type: none"> <li>- Wörter strukturieren und Möglichkeiten der Wortbildung kennen,</li> <li>- Wörter sammeln und ordnen,</li> <li>- sprachliche Operationen nutzen: umstellen, ersetzen, ergänzen, weglassen,</li> <li>- die Textproduktion und das Textverständnis durch die Anwendung von sprachlichen Operationen unterstützen,</li> <li>- mit Sprache experimentell und spielerisch umgehen.</li> </ul>
<b>grundlegende sprachliche Strukturen und Begriffe kennen und verwenden, z.B.</b>	<b>Wort</b>	<ul style="list-style-type: none"> <li>- Buchstabe, Laut, Selbstlaut, Mitlaut, Umlaut, Silbe, Alphabet</li> <li>- Wortfamilie, Wortstamm, Wortbaustein; Wortfeld; Wortart</li> <li>- Nomen: Einzahl, Mehrzahl, Fall, Geschlecht</li> <li>- Verb: Grundform, gebeugte Form</li> <li>- Zeitformen: Gegenwart, Vergangenheitsformen</li> <li>- Artikel: bestimmter Artikel, unbestimmter Artikel</li> <li>- Adjektiv: Grundform, Vergleichsstufen</li> <li>- Pronomen</li> <li>- andere Wörter (alle hier nicht kategorisierten Wörter gehören zu dieser Restkategorie)</li> </ul>
	<b>Satz</b>	<ul style="list-style-type: none"> <li>- Satzzeichen: Punkt, Komma, Fragezeichen, Ausrufezeichen,</li> <li>- Doppelpunkt, Redezeichen</li> <li>- Satzart: Aussage-, Frage-, Ausrufesatz</li> <li>- wörtliche Rede</li> <li>- Subjekt</li> <li>- Prädikat/Satzkern</li> <li>- Ergänzungen: Satzglied; einteilige, mehrteilige Ergänzung</li> <li>- Vergangenheit, Gegenwart, Zukunft (als Zeitstufen)</li> </ul>

Tabelle 4: Inhalte der bei VERA berücksichtigten Teilkompetenzen „An Wörtern, Sätzen, Texten arbeiten“ und „Grundlegende sprachliche Strukturen und Begriffe kennen und verwenden“

Die Aufgaben zu diesem Kompetenzbereich beziehen sich auf Semantik und Stil, Morphologie, Syntax sowie Regelwissen und Rechtschreibung. Betont werden die Alltagsnähe der Aufgaben und die (implizite) Berücksichtigung grammatikalischer Begrifflichkeiten bei deren Konstruktion. Die Aufgabenformate sind dabei breit gefächert und beinhalten geschlossene Formate wie Multiple-Choice, Ergänzungs-, Zuordnungs- und Unterstreichungsarbeiten genauso wie halboffene Formate (vgl. Projekt VERA, 2008; Isaac, Metzeld & Eichler, 2009).

### 2.4.1.2 Kompetenzskalen im Fach Mathematik bei VERA 3

In der Grundschule soll der Mathematikunterricht entsprechend den Vorstellungen, wie sie in den Bildungsstandards dargelegt werden, mathematische Alltagserfahrungen der Schülerinnen und Schüler aufnehmen und weiterentwickeln (KMK, 2004b). Aus ihnen sollen sich frühe Kompetenzen entfalten, die eine Grundlage für die Auseinandersetzung mit mathematischen Anforderungen in späteren Lebenssituationen (inklusive der Schule) bilden. Darüber hinaus betonen die Bildungsstandards auch die Bedeutung einer grundlegenden positiven Einstellung gegenüber der Mathematik, die bereits in der Grundschule bestmöglich gefördert werden soll.

Kompetenzen  
in der  
Primarstufen-  
mathematik

Die Bildungsstandards der Primarstufe für das Fach Mathematik beziehen sich nur indirekt auf Sachgebiete des Mathematikunterrichts der Grundschule (Arithmetik, Geometrie, Größen und Sachrechnen). Vielmehr werden allgemeine und inhaltsbezogene mathematische Kompetenzen abgegrenzt.

*Allgemeine mathematische Kompetenzen* beziehen sich auf die Art und Weise, wie Mathematik betrieben wird, und werden in der Auseinandersetzung mit Mathematik erworben. Die Ausprägung der jeweiligen Kompetenzen ist damit abhängig von der Quantität und Qualität der Unterrichtsgelegenheiten, in denen diese erprobt werden können. Die Bildungsstandards unterscheiden sechs Kompetenzbereiche (vgl. auch die Umschreibung der mathematischen Kompetenzen bei VERA 8 unten):

Allgemeine  
mathe-  
matische  
Kompetenzen

1. Technische Grundfertigkeiten
2. Problemlösen
3. Kommunizieren
4. Argumentieren
5. Modellieren
6. Darstellen

*Inhaltsbezogene mathematische Kompetenzen* beziehen sich dagegen über die Grenzen der (curricularen) Sachgebiete des Mathematikunterrichts hinweg auf grundlegende mathematische Konzepte, die in fünf mathematische Leitideen gegliedert sind:

Inhalts-  
bezogene  
mathe-  
matische  
Kompetenzen

1. Zahlen und Operationen
2. Raum und Form
3. Muster und Strukturen
4. Größen und Messen
5. Daten, Häufigkeit und Wahrscheinlichkeit

Bereits in der Primarstufen-Mathematik greifen die Bildungsstandards damit wesentliche Ideen von Freudenthal (1977; 1983) auf, wonach der verständnisvolle Umgang mit Mathematik und ihre flexible Anwendung in verschiedenen Kontexten im Zentrum mathematischen Denkens stehen. In der Primarstufe kann diese Entwicklung in Gang gesetzt werden, indem Alltagserfahrungen der Schülerinnen und Schüler aufgegriffen werden.

Das vom IQB entwickelte Kompetenzstufenmodell beinhaltet alle in den Bildungsstandards ausgewiesenen mathematischen Leitideen und ist in fünf Kompetenzstufen aufgebaut, die von der Beschreibung basaler Kompetenzen bis hin zum elaborierten Umgangs mit mathematischen Inhalten in der Primarstufe reichen (vgl. IQB, 2010; vgl. auch Reiss & Winkelmann 2009). Zentrale Kompetenzanforderungen dieser Stufen sind in Tabelle 5 (S. 20) dargestellt (IQB, 2010). Für die Vergleichsarbeiten in der Grundschule wurden in den letzten Jahren jeweils zwei oder drei mathematische Leitideen ausgewählt, für die eine kompetenzbezogene Rückmeldung erfolgte.

In den Anforderungsbereichen „Reproduzieren“, „Zusammenhänge erkennen“ und „Verallgemeinern und Reflektieren“ wird die (kognitive) Komplexität und Qualität der Kompetenzausprägung in den Bildungsstandards grob differenziert. Eine Anbindung an diese Anforderungsbereiche ist bei VERA allerdings bisher nicht vorgenommen worden.

Anforderungs-

Zumeist wird auf geschlossene Aufgabenformate zurückgegriffen, insbesondere auf Ergänzungs- und Multiple-Choice-Aufgaben, die adressatengerecht häufig durch mathematische Abbildungen als Stimulus eingeleitet werden.

Aufgaben-  
formate

### 2.4.2 Vergleichsarbeiten in der achten Jahrgangsstufe (VERA 8)

Das Institut zur Qualitätsentwicklung im Bildungswesen (IQB) hatte im Zuge der Normierung der Bildungsstandards eine umfangreiche Aufgabenentwicklung vorangetrieben, aus der seit 2009 auch die Aufgaben für die achte Jahrgangsstufe hervorgehen. Die Kompetenzen werden auf fünf gleich großen Stufen beschrieben, wobei die Zuordnung zu Stufe 2 im Sinne des Erreichens der Mindeststandards (vgl. Kapitel 3.2 im Studienbrief Bildungsmonitoring) diskutiert wird.

Aufgaben-  
entwicklung  
bei VERA 8

Um der relativ großen Leistungsstreuung in Deutschland gerecht zu werden, werden Aufgabensätze in drei Schwierigkeitsstufen (Testheftversionen) entwickelt (vgl. <http://www.iqb.hu-berlin.de/vera/faq>). Welche dieser Schwierigkeitsstufen in welcher Schulart eingesetzt wird, bestimmen die Länder. Zumeist erhalten aber alle Schülerinnen und Schüler einer Klasse dieselbe Testheftversion. Wie in Kapitel 2.3 erläutert, ist es möglich, Testergebnisse dieser unterschiedlich schwierigen Testhefte zu vergleichen und damit auch die erworbenen Kompetenzen in Relation zu setzen.

Aufgabenzusam-  
menstellungen in  
drei Schwierig-  
keitsstufen

Stufe	zentrale Anforderungsmerkmale
I a: Technische Grundlagen 1	Die Grundaufgaben des kleinen Einspluseins und Einmaleins werden beherrscht und genutzt, wenn die Aufgabenstellungen keine besonderen Schwierigkeiten aufweisen. Klar strukturierten Diagrammen, Schaubildern und Tabellen mit Bezug zur Lebenswirklichkeit können relevante Daten entnommen werden.
I b: Technische Grundlagen 2	Einfache mathematische Begriffe und Prozeduren sind bekannt und können in einem innermathematischen Kontext bzw. in einem aus dem Alltag vertrauten oder gut geübten Kontext korrekt reproduziert werden. Kleine Zahlen können in Bezug auf ihre Größe verglichen werden, Zahldarstellungen in Stellentafeln werden sicher gelesen. Auch die schwierigeren Einmaleinsaufgaben werden gelöst. Numerisches Wissen wird in einfachen Alltagssituationen angewendet. Einfache Reihungen werden erkannt und fortgesetzt. Insbesondere werden grundlegende Begriffe der ebenen Geometrie und gängige Repräsentanten standardisierter Einheiten richtig verwendet.
II: Einfache Anwendungen von Grundlagenwissen	Hier geht es um Routineprozeduren in einem klar strukturierten Kontext: Die Struktur des Dezimalsystems wird genutzt, Gesetzmäßigkeiten werden erkannt und bei der Fortsetzung einfacher Zahlenfolgen, beim strukturierten Zählen und systematischen Probieren berücksichtigt. Aufgaben zur Addition, Subtraktion und Multiplikation werden halbschriftlich und schriftlich durchgeführt, Überschlagsrechnungen werden durchgeführt. Insbesondere können in diesem Zusammenhang einfache Sachaufgaben gelöst werden. Aus dem Alltag vertraute proportionale Zuordnungen werden erkannt und angewendet. Bei einfachem Zahlenmaterial wird das Umwandeln von Größen in gegebene Einheiten auch bei gemischten Größenangaben durchgeführt. Grundbegriffe der räumlichen Geometrie werden korrekt verwendet, wenn diese einen Bezug zum Alltag haben. Räumliche Beziehungen werden zur Lösung einfacher Probleme genutzt. Wesentliche Grundbegriffe aus dem Umfeld von Zufall und Wahrscheinlichkeit werden korrekt verwendet („sicher“, „unmöglich“, „wahrscheinlich“).
III: Erkennen und Nutzen von Zusammenhängen	Potenzielle Zusammenhänge werden erkannt und in einem vertrauten (mathematischen und sachbezogenen) Kontext genutzt. Das erlernte Wissen kann auf dieser Stufe flexibel in unterschiedlichen Problemstellungen genutzt werden, die einem vertrauten Kontext zuzuordnen sind. Insbesondere wird mit Zahlen und Operationen im curricularen Umfang sicher umgegangen. Überschlagsrechnungen werden auch bei großen Zahlen sicher durchgeführt. Strukturelle Aspekte werden bei gut geübten Inhalten gesehen und können kommuniziert werden. Das betrifft auch Inhalte der Geometrie, wobei etwa zwischen verschiedenen Darstellungsformen einer Figur vermittelt werden kann. Einfache Sachsituationen werden modelliert und die damit verbundenen Problemstellungen gelöst. Daten und Informationen können in bekanntem Kontext flexibel dargestellt werden. Bei nicht allzu komplexen Zufallsexperimenten werden Gewinnchancen korrekt eingeschätzt und begründet.
IV: Sicheres und flexibles Anwenden von begrifflichem Wissen und Prozeduren in curricularem Umfang	Auch in einem wenig vertrauten Kontext wird mathematisches Wissen sicher angewendet. Eigene Vorgehensweisen werden korrekt beschrieben, die Lösungswege anderer Kinder werden verstanden und reflektiert. Das Rechnen wird im curricularen Umfang in allen Varianten sicher beherrscht. Begriffe der ebenen und räumlichen Geometrie werden flexibel verwendet. Zahldarstellungen in Stellenwerttafeln können auch bei sehr großen Zahlen nach Vorschrift selbstständig manipuliert und systematisch verändert werden. Das Rechnen mit Größen ist sicher und flexibel und umfasst insbesondere Näherungsrechnungen und Überschlagsrechnungen. Informationen aus unterschiedlichen Quellen können in einen Zusammenhang gestellt und in Modellierungsaufgaben selbstständig verwendet und manipuliert werden.
V: Modellierung komplexer Probleme unter selbstständiger Entwicklung geeigneter Strategien	Mathematische Problemstellungen werden auch in einem unbekannten Kontext angemessen, sicher und flexibel bearbeitet. Dabei werden geeignete Strategien, sinnvolle Bewertungen oder Verallgemeinerungen auf hohem Niveau geleistet. Umfangreiches curricular verankertes Wissen wird in ungewohnten Situationen flexibel genutzt. Das Vorgehen kann sicher und nachvollziehbar kommuniziert und begründet werden. Komplexe Sachsituationen werden modelliert und bearbeitet, wobei besondere Schwierigkeiten wie die Verwendung von Tabellen, der Umgang mit zusammengesetzten Größen oder das Rechnen mit Zahlen in Kommaschreibweise auftreten können. Es können auch ungewohnte funktionale Zusammenhänge analysiert und genutzt werden. Die Lösung von Aufgaben kann ein hohes Maß an räumlichem Denken oder entsprechende analytische Fähigkeiten voraussetzen.

Tabelle 5: Zentrale Anforderungsmerkmale der Kompetenzstufen im Fach Mathematik aus der VERA 3-Erhebung 2010 (IQB, 2010)

Die folgende Darstellung der Kompetenzstufenmodelle für Deutsch, Englisch und Mathematik kann durch Aufgabenbeispiele illustriert werden (<http://www.iqb.hu-berlin.de/vera/aufgaben>).

#### 2.4.2.1 Kompetenzskalen im Fach Deutsch bei VERA 8

Kompetenz in der deutschen Sprache geht nach den Vorstellungen der KMK weit über eine reine Verstehens- und Verständigungskompetenz hinaus (KMK, 2003a; 2004c). Die in den Bildungsstandards beschriebenen Erwartungen sollen vielmehr als Voraussetzungen dafür dienen, dass Schülerinnen und Schüler Orientierungs- und Handlungswissen gewinnen, etwa wenn es darum geht, „kritische Distanz zwischen Lebenswirklichkeit und den in Literatur und Medien dargestellten virtuellen Welten“ zu entwickeln, sich die „Bedeutung des Reichtums kultureller, sprachlicher, literarischer und medialer Vielfalt“ zu vergegenwärtigen oder im direkten Austausch und in der Auseinandersetzung mit kulturellen Traditionen Fremdverstehen und Toleranz zu fördern. Die Vorlage der KMK für die Entwicklung von Kompetenzstufenmodellen zur Untersuchung des Lernstands im Fach Deutsch war also umfangreich.

Kompetenz-  
skalen im  
Fach Deutsch  
bei VERA 8

Im Fach Deutsch gingen weitreichende Überlegungen zur Aufgabenentwicklung und zu den Kompetenzstufenmodellen aus dem Projekt „Evaluation der Standards Deutsch für die Sekundarstufe I“ (ESDeS I; Federführung: Prof. Dr. Bremerich-Vos, Universität Duisburg-Essen) hervor. Darüber hinaus lagen Erfahrungen aus den Large-Scale-Assessments des Bildungsmonitorings, insbesondere aus PISA (Kompetenzbereich Lesen, vor allem PISA 2000; Artelt, Stanat, Schneider & Schiefele, 2001) und DESI (DESI-Konsortium, 2008) vor. Aktuell liegen für den mittleren Schulabschluss Kompetenzstufenmodelle für folgende Kompetenzbereiche vor (vgl. die Webseite des IQB: [http://www.iqb.hu-berlin.de/bista?reg=r\\_4](http://www.iqb.hu-berlin.de/bista?reg=r_4); 22.06.2011):

- Lesen
- Orthografie
- Sprechen und Zuhören (hier zunächst allerdings nur die Subkomponente „verstehend zuhören“).

Die konstruierten Aufgaben „tragen den aktuellen Erkenntnissen der Fachdidaktik und der Psychometrie Rechnung, um einerseits der Bedeutung des Deutschunterrichts hinsichtlich seines Beitrags zur sprachlichen, literarischen und medialen Bildung der Schülerinnen und Schüler gerecht zu werden und andererseits empirisch verlässliche Messinstrumente zu erhalten“ (IQB, ohne Jahr). Dies wird hier beispielhaft für den Kompetenzbereich Lesen beschrieben (vgl. IQB, 2009).

#### Beispiel: Kompetenzbeschreibungen im Lesen – Komponente „Lesen – mit Texten und Medien umgehen“

Leseverstehen beschreibt die Fähigkeit, als Schrift- oder Bildzeichen kodierte Informationen in schriftlichen Dokumenten zu verstehen (Schnotz & Dutke, 2004). Diese Informationen können als Texte, aber auch in Form von Bildern, Diagrammen oder Tabellen dargestellt sein. Leseverstehen beinhaltet kognitive Prozesse, bei der die Lesenden aktiv eine mentale Repräsentation der Textinhalte vornehmen und ihnen Bedeutung verleihen. Zudem ist für die Ausprägung von Lesekompetenz die Motivation zum Lesen relevant, von der angenommen wird, dass sie über die Lesemenge und den Einsatz von Lesestrategien die Kompetenzausprägung beeinflusst (Möller & Schiefele, 2004).

Beispiel  
Kompetenz-  
bereich  
„Lesen“

In den Bildungsstandards für das Fach Deutsch besitzt die Lesekompetenz - zweifellos auch unter dem Eindruck von PISA - eine hohe Bedeutung: Lesekompetenz wird darin als Schlüsselkompetenz zur Teilhabe am gesellschaftlichen Leben angesehen. Einige der dort beschriebenen Komponenten (z.B. Anschlusskommunikation) sind diagnostisch nur schwer zu erfassen, insbesondere im Rahmen von groß angelegten Tests im Klassenverband.

Die bisher entstandenen Kompetenzstufenmodelle berücksichtigen vornehmlich das Leseverstehen bei literarischen Texten sowie Sach- und Gebrauchstexten (vgl. die Kompetenzanforderungen in Tabelle 6, S.22). Es wurde hier also weitestgehend den konzeptionellen Vorarbeiten aus Studien des Bildungsmonitoring wie PISA oder IGLU (PIRLS) gefolgt, bei denen ebenfalls nach dem Textgenre differenziert wird. Daneben sind Leseprozesse (Informationen ermitteln vs. textbezogenes Interpretieren vs. Reflektieren und Bewerten) und das Textformat (kontinuierliche vs. diskontinuierliche Texte) als relevante differenzierende Merkmale der Lesekompetenz erfasst worden. Alle konstruierten Aufgaben lassen sich dennoch auf einer gemeinsamen Kompetenzskala abbilden, da Schülerinnen und Schülern in diesen verschiedenen Merkmalsbereichen sehr ähnliche Ergebnisse erbringen.

Bei der Erfassung von Lesekompetenz in den Vergleichsarbeiten stehen die folgenden Aspekte im Vordergrund:

- wesentliche Elemente eines Textes erfassen
- wesentliche Fachbegriffe zur Erschließung von Literatur kennen und anwenden
- eigene Deutungen des Textes entwickeln
- Informationen zielgerichtet entnehmen, ordnen (...)
- Medien verstehen und nutzen
- zwischen eigener Wirklichkeit und virtuellen Welten in Medien unterscheiden.

**Grundlegende Verfahren**

- Lesetechniken und Strategien zum Leseverstehen kennen und anwenden
- über grundlegende Lesefertigkeiten verfügen: flüssig, sinnbezogen, überfliegend, selektiv
- die eigenen Leseziele kennen
- Vorwissen und neue Informationen unterscheiden
- Wortbedeutungen klären
- Lesehilfen nutzen: z.B. Textsorte, Aufbau, Überschrift, Illustration, Layout
- Verfahren zur Textstrukturierung kennen und nutzen: Inhalte zusammenfassen, Zwischenüberschriften formulieren, wesentliche Textstellen kennzeichnen, Bezüge zwischen Textstellen herstellen, Fragen aus dem Text ableiten und beantworten
- Verfahren zur Textaufnahme kennen und nutzen: Aussagen erklären, Stichwörter formulieren, Texte und Textabschnitte zusammenfassen

**Literarische Texte verstehen und nutzen**

- aktuelle und klassische Werke der Jugendliteratur und altersangemessene Texte bedeutender Autorinnen und Autoren kennen
- epische, lyrische, dramatische Texte unterscheiden und wesentliche Merkmale kennen, insbesondere epische Kleinformen, Erzählung, Kurzgeschichte, Gedichte
- an einem repräsentativen Beispiel Zusammenhänge zwischen Text, Entstehungszeit und Leben des Autors/der Autorin herstellen
- zentrale Aussagen erschließen
- wesentliche Elemente eines Textes erfassen: Figuren, Raum- und Zeitdarstellung, Konfliktverlauf
- Handlung und Verhaltensweisen beschreiben und werten
- wesentliche Fachbegriffe zur Erschließung von Literatur kennen und anwenden: Autor, Erzähler, Monolog, Dialog, Reim
- grundlegende Gestaltungsmittel erkennen und ihre Wirkungen einschätzen: z.B. Wortwahl, Wiederholung, sprachliche Bilder
- untersuchende und produktive Methoden kennen und anwenden: z.B. Texte vergleichen, weiterschreiben, Paralleltext verfassen, szenische Umsetzung
- eigene Deutungen des Textes entwickeln, mit anderen darüber sprechen und am Text belegen

**Sach- und Gebrauchstexte verstehen und nutzen**

- verschiedene Textfunktionen und Textsorten unterscheiden: informieren (z.B. Lexikontext), appellieren (z.B. Werbetext), regulieren (z.B. Jugendschutzgesetz, Arbeitsvertrag), instruieren (z.B. Bedienungsanleitung)
- Informationen zielgerichtet entnehmen, ordnen, prüfen und ergänzen
- nichtlineare Texte (auch im Zusammenhang mit linearen Texten) auswerten: z.B. Schaubilder,
- Intention(en) eines Textes erkennen
- aus Sach- und Gebrauchstexten begründete Schlussfolgerungen ziehen
- Information und Wertung in Texten unterscheiden: z.B. in Zeitungen

**Medien verstehen und nutzen**

- Informations- und Unterhaltungsfunktion unterscheiden: z.B. im Internet
- wesentliche Darstellungsmittel eines Mediums und deren Wirkungen kennen und einschätzen
- Intentionen und Wirkungen ausgewählter Medieninhalte erkennen und bewerten: z. B. Fernsehserie
- Lebenswirklichkeit von Realitätsdarstellungen und der Darstellung fiktionaler Welten in Medien unterscheiden
- Informationen zu einem Thema/Problem in unterschiedlichen Medien suchen, vergleichen, auswählen und bewerten
- Medien für die eigene Produktion kreativ nutzen

**Methoden und Arbeitstechniken**

- mit Nachschlagewerken umgehen können
- recherchieren
- zitieren, Quellen angeben
- Wesentliches markieren
- Stichwörter formulieren
- Texte gliedern und Teilüberschriften finden
- Inhalte mit eigenen Worten zusammenfassend wiedergeben
- Arbeitsergebnisse zielgerichtet und sachbezogen präsentieren z.B. mit Folie, Plakat, PC

Tabelle 6: Teilkompetenzen des Kompetenzbereichs „Lesen – mit Texten und Medien umgehen“ im Fach Deutsch in den Bildungsstandards für den Hauptschulabschluss



Überwiegend wird dabei auf authentische Texte zurückgegriffen, die sich in Länge und Komplexität deutlich unterscheiden können. Als Aufgabenformate kommen zumeist Multiple-Choice-, Wahr/Falsch- und Zuordnungsaufgaben zum Einsatz, teilweise sind Kurzantworten zu geben. Alle Aufgaben werden dichotom ausgewertet, es gibt also keine „Zwischen-“ oder „Teillösungen“, und auch nicht bearbeitete Items werden als „falsch“ gewertet. Diese Art der Auswertung unterscheidet sich damit beispielsweise von differenzierteren Bewertungsschemata in Klassenarbeiten, bei deren Korrektur das Prinzip gilt, jede erbrachte Leistung durch Teilpunkte zu würdigen. Vergleichsarbeiten sind dagegen so konzipiert, dass mit dem Ergebnis ausgelotet werden kann, auf welcher Kompetenzstufe sich die Schülerinnen und Schüler mit einer gewissen Wahrscheinlichkeit befinden. Damit Schülern z. B. eine hohe Kompetenzstufe zugewiesen werden kann, müssen sie in der Lage sein, mit einer hinreichenden Sicherheit auch schwierige, komplexe Aufgaben vollständig zu lösen. Um bei diesen Aufgaben zu Teillösungen zu gelangen, würde eine geringere Kompetenz genügen, welche sich aber bereits durch die Bearbeitung von leichten Aufgaben nachweisen lässt.

Aufgaben-  
formate Lese-  
verstehen

#### 2.4.2.2 Kompetenzskalen im Fach Englisch bei VERA 8

Tenorth (2001) stellte fest, dass der Beherrschung der englischen Sprache als lingua franca eine besondere Bedeutung unter den Neu- und Alt Sprachen zukommt. Sprachliche Kompetenz im Englischen ist längst eine „notwendige Voraussetzung für die Teilnahme an der Leistungsgesellschaft“. So verwundert es nicht, dass sich die KMK mit dem Auftrag für die Schulleistungsstudie DESI (Deutsch-Englisch-Schülerleistungen-International; DESI-Konsortium, 2008; vgl. Studienbrief 3) bemühte, eine Lücke in der Gesamtstrategie zum Bildungsmonitoring zu schließen, und Vergleichsarbeiten in Englisch als erster Fremdsprache der meisten Schülerinnen und Schüler – neben Deutsch und Mathematik – in nahezu allen Bundesländern implementiert wurden.

Kompetenz-  
skalen im Fach  
Englisch  
bei VERA 8

Die Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) greifen auf den Gemeinsamen Europäischen Referenzrahmen für Sprachen (GER; Europarat, 2001; vgl. Tabelle 7) zurück, in dem produktive und rezeptive Kompetenzen der Fremdsprachenverwendung auf verschiedenen Dimensionen definiert sind. Diese Kompetenzen sollen die Handlungsfähigkeit von Personen im öffentlichen, privaten und beruflichen Leben gewährleisten.

Bildungs-  
standards  
in der ersten  
Fremdsprache

Für das Fach Englisch definieren die nationalen Bildungsstandards Anforderungen in Bezug auf funktionale kommunikative, interkulturelle und methodische Kompetenzen (Abbildung 7, S. 24; vgl. KMK, 2003b; 2004d). Allerdings lassen sich vor allem kommunikative Kompetenzen in der englischen Sprache im Rahmen von Vergleichsarbeiten sinnvoll erfassen, sodass sich die bisher entwickelten Kompetenzstufenmodelle auf die Kompetenzbereiche Hörverstehen und Leseverstehen beschränken ([http://www.iqb.hu-berlin.de/bista?reg=r\\_4](http://www.iqb.hu-berlin.de/bista?reg=r_4); 22.06.2011).

Leistungs-  
niveaus  
des GER

Im GER sind in Form von Deskriptoren erwartbare Leistungen definiert, die insgesamt sechs Leistungsniveaus (A1, A2, B1, B2, C1, C2) unterscheiden. Durch das IQB wurden Testaufgaben entwickelt, welche die Niveaus A1 bis C1 abdecken, sodass sich in den Modellen also insgesamt fünf Kompetenzstufen abgrenzen lassen. Das Niveau C2 wurde hingegen ausgelassen, da es sich hierbei um Anforderungen handelt, die in der Regel erst in der Sekundarstufe II Berücksichtigung finden. Die Bildungsstandards für den Hauptschulabschluss orientieren sich im Kern an den Niveaus A2 und B1, die Bildungsstandards für den Mittleren Schulabschluss an den Niveaus B1 und B2. Tabelle 8 (S. 25) zeigt beispielhaft die Zuordnung von Kompetenzerwartungen zu den GER-Niveaus anhand des Kompetenzbereichs Hörverstehen. Die Aufgaben der Vergleichsarbeiten beinhalten ebenfalls eine Zuschreibung zu den Niveaustufen des GER.

Inhalte, Schwierigkeitsmerkmale und Operationalisierung der Standards für das Fach Englisch sowie Implikationen für Vergleichsarbeiten werden im Folgenden am Beispiel des Kompetenzbereiches „Hörverstehen“ dargestellt.

deutscher Schulabschluss	GER-Niveau	Spezifikation des Niveaus
Hauptschulabschluss	A:	Elementare Sprachverwendung A1: Breakthrough A2: Waystage
mittlerer Bildungsabschluss	B:	Selbstständige Sprachverwendung B1: Threshold B2: Vantage
Abitur	C:	Kompetente Sprachverwendung C1: Effective Operational Proficiency C2: Mastery

Tabelle 7: Niveaus des gemeinsamen europäischen Referenzrahmens (GER)

#### Beispiel: Kompetenzbeschreibungen Englisch, Kompetenz „Hörverstehen“, mittlerer Bildungsabschluss

Bei der Erfassung von Hörverstehen wird der Stimulus „in Echtzeit“ akustisch dargeboten. Vor allem Besonderheiten in der sprachlichen Darbietung wie Dialekte und die Flüchtigkeit der Stimulusdarbietung machen dabei die Schwierigkeit des Verstehensprozesses aus (siehe unten). Die kognitiven Prozesse beim Hörverstehen reichen von der akustischen Wahrnehmung kurzzeitig dargebotener Signale bis zur Sinngebung der akustischen Informationen durch die Integration von sprachlichem Wissen sowie Sach- und Weltwissen.

Beispiel  
Kompetenz-  
bereich  
„Hörver-  
stehen“

Hörverstehen ist bereits in der Erstsprache ein komplexer Vorgang; beim Erlernen einer Fremdsprache werden jedoch besonders hohe Anforderungen an das Hörverstehen gestellt. Dies liegt zum einen daran, dass Verstehenslücken

cken beim Hörverstehen aufgrund der kurzzeitigen, flüchtigen Darbietung des akustischen Reizes nur noch sehr schwer gefüllt werden können. Zum anderen setzt Hörverstehen auch Vorwissen voraus. Dies bezieht sich nicht nur auf Vokabeln, sondern insbesondere auch auf Sinnzusammenhänge, die nur durch Wissen über Inhalte und kulturelle Besonderheiten hergestellt werden können. Fehlt dieses Wissen, so können Missverständnisse entstehen oder Texte nicht verstanden werden.

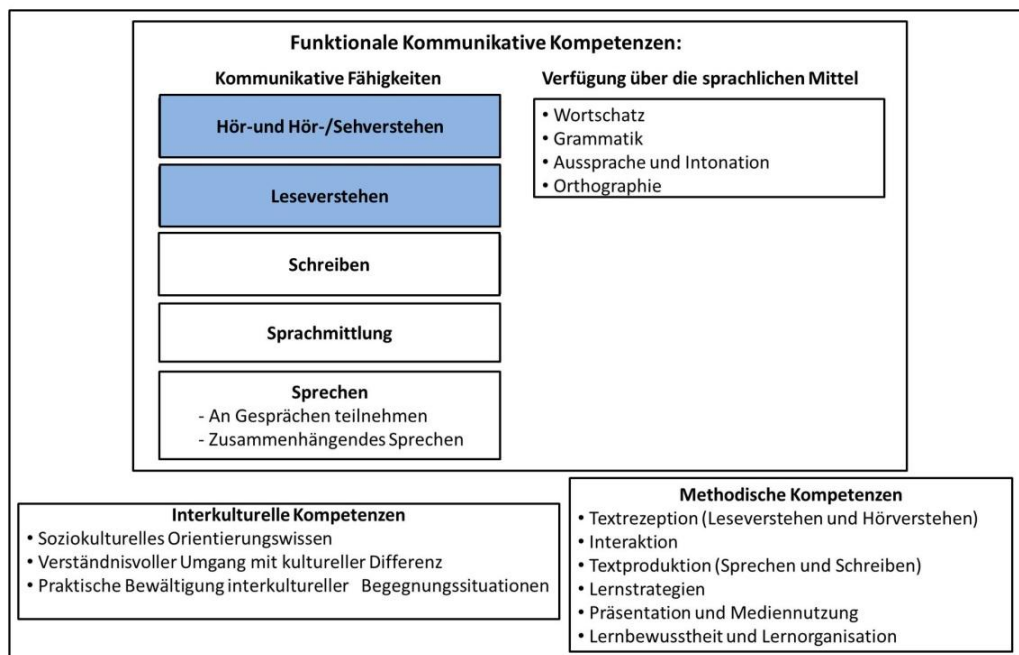


Abbildung 7: Funktional kommunikative, interkulturelle und methodische Kompetenzen der Bildungsstandards für den Hauptschul- und mittleren Bildungsabschluss in der ersten Fremdsprache

Hinweis: Blau unterlegt sind jene Kompetenzbereiche, für die durch das IQB Kompetenzstufenmodelle entwickelt wurden ([http://www.iqb.hu-berlin.de/bista?reg=r\\_4](http://www.iqb.hu-berlin.de/bista?reg=r_4); 22.06.2011).

Aber auch die Eigenschaften der Hörtexte und die Formulierung der Höraufgabe beeinflussen den Schwierigkeitsgrad des Hörverstehens. So haben sich als schwierigkeitsbestimmende Merkmale des Textes u. A. herausgestellt: die Länge des Textes, die Sprechgeschwindigkeit, die Anzahl und Aussprache der Sprecherinnen und Sprecher, Eigenheiten der verwendeten Sprache (z.B. Dialekte), die Textstruktur (Gliederung des Textes) und die lexikalische und grammatische Komplexität (Häufigkeit oder Abstraktionsgrad der Begriffe, Komplexität der Satzstrukturen).

Die Bildungsstandards betonen die Bedeutung von Hörverstehen als Teil kommunikativer Kompetenzen, die es ermöglichen, sich in der Fremdsprache „in Alltagssituationen und über lebenspraktische Angelegenheiten“ zu verständigen (KMK, 2003b). Hörverstehen besitzt aber auch Relevanz für die spätere berufliche Tätigkeit, etwa bei Telefonaten oder Bewerbungen im Ausland. Die Kompetenzbereiche des Hörverstehens beinhalten in den Bildungsstandards die Rezeption direkter Kommunikation zwischen Sprechern als auch mediale Kommunikation.

In der Formulierung der Bildungsstandards heißt es (KMK, 2003b; jeweiliges GER-Niveau in Klammern): „Die Schülerinnen und Schüler können unkomplizierte Sachinformationen über gewöhnliche alltags- oder berufsbezogene Themen verstehen und dabei die Hauptaussagen und Einzelinformationen erkennen, wenn in deutlich artikulierter Standardsprache gesprochen wird (B1+).“

Die Schülerinnen und Schüler können

- im Allgemeinen den Hauptpunkten von längeren Gesprächen folgen, die in ihrer Gegenwart geführt werden (B1),
- Vorträge verstehen, wenn die Thematik vertraut und die Darstellung unkompliziert und klar strukturiert ist (B1+),
- Ankündigungen und Mitteilungen zu konkreten Themen verstehen, die in normaler Geschwindigkeit in Standardsprache gesprochen werden (B2),
- vielen Filmen folgen, deren Handlung im Wesentlichen durch Bild und Aktion getragen wird (B1),
- den Informationsgehalt der meisten Rundfunksendungen und Tonaufnahmen über Themen von persönlichem Interesse verstehen (B1+),
- das Wesentliche in vielen Fernsehsendungen zu Themen von persönlichem Interesse, z. B. Interviews, kurze Vorträge oder Nachrichtensendungen verstehen (B1+).“

Die Aufgaben zur Erfassung von Hörverstehen werden dahingehend konstruiert, dass sie möglichst authentische Texte aus englischsprachigen Ländern und unterschiedliche Textarten aufgreifen sowie eine Progression in den Leis-

tungsanforderungen beinhalten. Anwendung finden vor allem Aufgabenformate wie Multiple Choice oder Multiple Matching, die eigene Textproduktion auf ein Mindestmaß reduzieren. Bei der Auswertung spielen dementsprechend orthographische und grammatikalische Fehler nur eine untergeordnete Rolle.

<b>A Elementare Sprachverwendung</b>	
<b>A1</b>	Kann verstehen, wenn sehr langsam und sorgfältig gesprochen wird und wenn lange Pausen Zeit lassen, den Sinn zu erfassen.
<b>A2</b>	Versteht genug, um Bedürfnisse konkreter Art befriedigen zu können, sofern deutlich und langsam gesprochen wird. Kann Wendungen und Wörter verstehen, wenn es um Dinge von ganz unmittelbarer Bedeutung geht (z. B. ganz grundlegende Informationen zu Person, Familie, Einkaufen, Arbeit, nähere Umgebung) sofern deutlich und langsam gesprochen wird.
<b>B Selbständige Sprachverwendung</b>	
<b>B1</b>	Kann unkomplizierte Sachinformationen über gewöhnliche alltags- oder berufsbezogene Themen verstehen und dabei die Hauptaussagen und Einzelinformationen erkennen, sofern klar artikuliert und mit vertrautem Akzent gesprochen wird. Kann die Hauptpunkte verstehen, wenn in deutlich artikulierter Standardsprache über vertraute Dinge gesprochen wird, denen man normalerweise bei der Arbeit, in der Ausbildung oder der Freizeit begegnet; kann auch kurze Erzählungen verstehen.
<b>B2</b>	Kann im direkten Kontakt und in den Medien gesprochene Standardsprache verstehen, wenn es um vertraute oder auch um weniger vertraute Themen geht, wie man ihnen normalerweise im privaten, gesellschaftlichen, beruflichen Leben oder in der Ausbildung begegnet. Nur extreme Hintergrundgeräusche, unangemessene Diskursstrukturen oder starke Idiomatik beeinträchtigen das Verständnis. Kann die Hauptaussagen von inhaltlich und sprachlich komplexen Redebeiträgen zu konkreten und abstrakten Themen verstehen, wenn Standardsprache gesprochen wird; versteht auch Fachdiskussionen im eigenen Spezialgebiet. Kann längeren Redebeiträgen und komplexer Argumentation folgen, sofern die Thematik einigermaßen vertraut ist und der Rede- oder Gesprächsverlauf durch explizite Signale gekennzeichnet ist.
<b>C Kompetente Sprachverwendung</b>	
<b>C1</b>	Kann genug verstehen, um längeren Redebeiträgen über nicht vertraute abstrakte und komplexe Themen zu folgen, wenn auch gelegentlich Details bestätigt werden müssen, insbesondere bei fremdem Akzent. Kann ein breites Spektrum idiomatischer Wendungen und umgangssprachlicher Ausdrucksformen verstehen und Registerwechsel richtig beurteilen. Kann längeren Reden und Gesprächen folgen, auch wenn diese nicht klar strukturiert sind und wenn Zusammenhänge nicht explizit ausgedrückt sind.
<b>C2</b>	Hat keinerlei Schwierigkeiten, alle Arten gesprochener Sprache zu verstehen, sei dies live oder in den Medien, und zwar auch wenn schnell gesprochen wird, wie Muttersprachler dies tun.

Tabelle 8: Deskriptoren des GER für die Niveaus A1 bis C2 im Kompetenzbereich „Hörverstehen“ in der ersten Fremdsprache

#### 2.4.2.3 Kompetenzskalen im Fach Mathematik bei VERA 8

Bei der Entwicklung von Aufgaben für das Fach Mathematik konnten die Beteiligten auf eine weitreichende Erfahrung aus den PISA-Studien (insbesondere PISA 2003; vgl. Studienbrief 3) zurückgreifen. Im Kontext der nationalen Ergänzungen der PISA-Studien in Deutschland waren eigene Aufgaben entwickelt worden, die der Aufgabenentwicklung für die nationalen Bildungsstandards im Fach Mathematik prägend vorausgingen. Charakteristisch für die Erfassung von mathematical literacy bei PISA ist die inhaltliche Differenzierung der Aufgaben nach allgemeinen mathematischen Kompetenzen und Leitideen (vgl. auch das Kompetenzstufenmodell für die Primarstufe oben). Für die Entwicklung der Aufgaben zur Überprüfung und Normierung der Bildungsstandards (KMK, 2003c; 2004e) wurde als zusätzliche Dimension der mathematische Anforderungsbereich spezifiziert. Diese Dimension war mehr oder weniger explizit ebenfalls bei PISA bereits herangezogen worden.

Die Aufgaben, welche die Bildungsstandards operationalisieren, füllen somit einen dreidimensionalen Raum, der fünf Leitideen, sechs Kompetenzen und drei Anforderungsbereiche abbildet (Abbildung 8, S. 26; vgl. Blum, Drüke-Noe, Hartung & Köller, 2006).

Das Konzept der Leitideen nimmt Überlegungen der Freudenthal-Didaktik (Freudenthal, 1977/1983) auf. Darin werden Phänomene einer mathematischen Betrachtungsweise beschrieben, welche den Stoffgebieten der Schulmathematik (Algebra, Arithmetik, Geometrie und Stochastik) vorausgehen und konzeptionell von diesen abzugrenzen sind. Die fünf *Leitideen* umfassen:

- Zahl,
- Messen,
- Raum und Form,
- Funktionaler Zusammenhang,
- Daten und Zufall.



## Anforderungsbereiche

In den *Anforderungsbereichen* wird die Komplexität einer Aufgabe berücksichtigt. Diese ist nicht direkt auf die empirische Schwierigkeit übertragbar, spiegelt sich in der Regel darin aber wider. In den Bildungsstandards werden drei nach zunehmender Komplexität geordnete Anforderungsbereiche (AB) definiert, deren Übergänge allerdings fließend sind:

1. Reproduzieren,
2. Zusammenhänge herstellen,
3. Verallgemeinern und reflektieren.

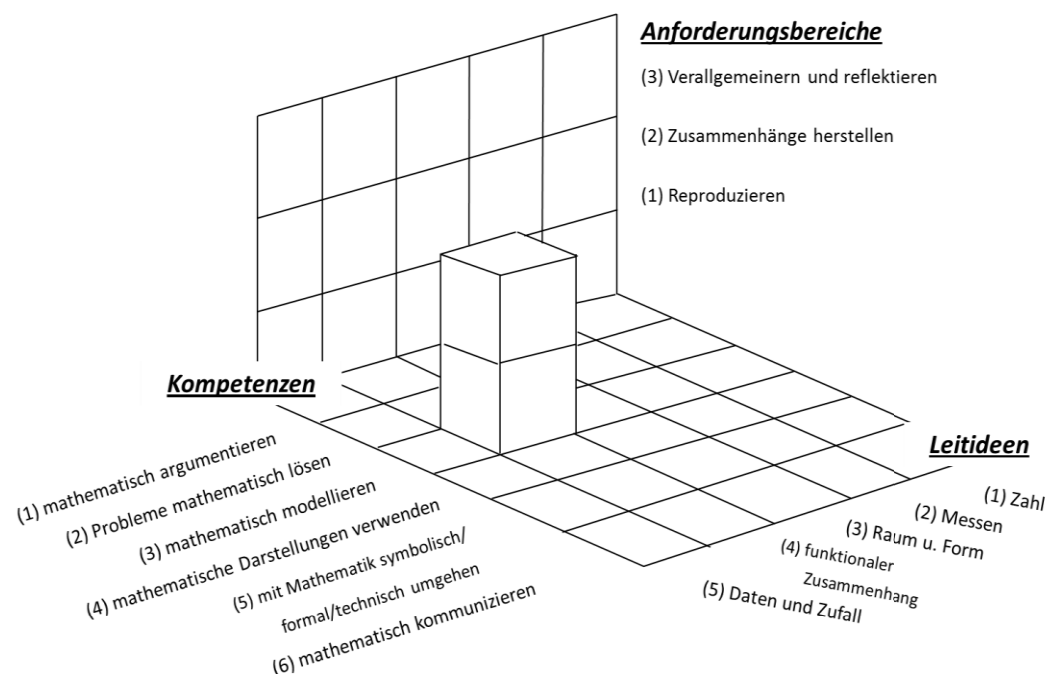


Abbildung 8: Dimensionen mathematischer Kompetenzen entsprechend der Bildungsstandards für den Hauptschul- und mittleren Bildungsabschluss im Fach Mathematik

## sechs mathematische Kompetenzen

Als mathematische *Kompetenzen* werden schließlich „zentrale Aspekte mathematischen Arbeitens“ (Blum, Drüke-Noe, Hartung & Köller, 2006) differenziert. Diese Kompetenzen stehen im Zentrum der Anforderungen aus den Bildungsstandards und beschreiben pragmatisch formuliert, welche Aktivitäten Schülerinnen und Schüler beim Bearbeiten von Mathematikaufgaben beherrschen sollen. Obwohl sich die Kompetenzen am Unterricht orientieren, stellen sie doch prinzipielle mathematische Arbeitsweisen dar, die auch außerhalb der Schule das „Mathematik betreiben“ charakterisieren. Die sechs mathematischen Kompetenzen (K) beinhalten:

**Mathematisch Argumentieren:** Mathematisches Argumentieren beinhaltet die Formulierung logischer Argumentationsketten sowie deren Verständnis und Bewertung. Im Zentrum steht die Rechtfertigung von Behauptungen auf Basis fundamentaler mathematischer Gesetze und Konventionen, wobei diese Rechtfertigung von einfachen Plausibilitätsüberlegungen bis zu strengen Beweisen reichen kann.

**Probleme mathematisch lösen:** Mathematische Probleme stellen Aufgaben dar, bei denen ein unmittelbarer Lösungsweg nicht zu erkennen ist und stattdessen unter Nutzung entsprechender Strategien (z.B. heuristischer Prinzipien oder Hilfsmittel) entwickelt werden muss. Darüber hinaus ist auch die Reflexion über entsprechende Strategien Teil dieser Kompetenz.

**Mathematisch modellieren:** Beim Modellieren steht die vereinfachte Abbildung der Realität, entweder realer Phänomene („deskriptive Modelle“) oder realer Sachverhalte („normative Modelle“), im Mittelpunkt. Ziel ist es, Sachverhalte einer Bearbeitung zugänglich zu machen, was Verständnis und Strukturierung der Realität voraussetzt und die Anforderung beinhaltet, eine Verbindung zwischen außermathematischem Kontext und innermathematischem Inhalt zu schaffen.

**Mathematische Darstellungen verwenden:** In diesem Kompetenzbereich werden Aspekte zusammengefasst, bei der die eigenständige Erzeugung und Veränderung sowie der Umgang mit und das Verständnis von mathematischen Darstellungen im Mittelpunkt stehen. Mathematische Darstellungen sind dadurch gekennzeichnet, dass diese Darstellungen (Abbildungen, Diagramme, etc.) explizit mathematische Informationen beinhalten und nicht bloß der Illustration oder Motivation dienen.

**Mit Mathematik symbolisch, formal und technisch umgehen:** Der fünfte Kompetenzbereich beinhaltet sowohl den Gebrauch mathematischer Fakten („Wissen, dass...“) als auch mathematischer Fertigkeiten („Wissen, wie...“). Dies bezieht sich auf mathematische Definitionen, Regeln, Algorithmen oder Formeln, welche bekannt sein und angewendet

werden sollen. Darüber hinaus wird hier auch formales Arbeiten mit Variablen, Termen, Gleichungen oder Funktionen und die Einhaltung einer bestimmten Schrittfolge bei der Aufgabenlösung erfasst. Auch der Umgang mit Hilfsmitteln, wie Formelsammlung oder Taschenrechner, wird berücksichtigt.

*Mathematisch kommunizieren:* In diesem Kompetenzbereich wird das Verstehen von mathematischen Texten oder mündlichen Beiträgen zur Mathematik berücksichtigt, außerdem verständliches schriftliches oder mündliches Präsentieren mathematischer Inhalte. Im Gegensatz zum Argumentieren lässt das Kommunizieren einen externen Adressatenbezug erwarten, sodass sprachliche Erläuterungen ins Gewicht fallen.

Für die Vergleichsarbeiten wird den Ländern derzeit eine repräsentative Aufgabenmenge zu allen Kompetenzen und Leitideen bereitgestellt. Die Aufgaben setzen sich aus einem Stimulus (Text oder Abbildung) und mehreren Items zusammen, die weitestgehend unabhängig voneinander lösbar sind. Als Aufgabenformate kommen Multiple Choice, Ergänzungsaufgaben oder andere Formate mit Kurzantworten in Frage. Teilweise wird die ausführliche Darlegung des Lösungswegs erwartet. Ungewöhnlich für viele Mathematiklehrerinnen und -lehrer ist das an Instrumenten des Bildungsmonitoring angelehnte Auswertungsschema der Aufgaben, welches nur die Optionen „Aufgabe gelöst“/„Aufgabe nicht gelöst“ vorsieht (vgl. S. 23 oben).

#### 2.4.3 Weiterführende Literatur

Informationen über die Vergleichsarbeiten in der dritten Jahrgangsstufe sind den Internetseiten des IQB (<http://www.iqb.hu-berlin.de/>) und der VERA-Projektgruppe (<http://www.uni-landau.de/vera/>) sowie den Internetseiten der jeweiligen Ministerien und Landesinstitute der Länder zu entnehmen. Bremerich-Vos, Granzer, Behrens & Köller (2009) liefern eine Beschreibung der Bildungsstandards in der Primarstufe für das Fach Deutsch. Das Analogon für das Fach Mathematik stammt von Walther, van den Heuvel-Panhuizen, Granzer & Köller (2007).

Die aktuellen Informationen über die Inhalte der Vergleichsarbeiten in der achten Jahrgangsstufe sind ebenfalls auf den Internetseiten des IQB bzw. den Webseiten der Ministerien und Landesinstitute zu finden. Eine detaillierte Beschreibung der Kompetenzbereiche in den Bildungsstandards in der Sekundarstufe I ist von Bremerich-Vos, Granzer und Köller (2008) vorgelegt worden. Informationen zur Normierung der Bildungsstandards in Englisch sind dem technical report<sup>10</sup> des IQB zu entnehmen (Rupp, Vock, Harsch & Köller, 2008). Konzept und erste Ergebnisse zur Normierung in Französisch – hier wie gesagt nicht vorgestellt – stellen Tesch, Leupold & Köller (2008) dar. Das Kompetenzstufenmodell und weiterführende Informationen zu den Bildungsstandards im Fach Mathematik erläutern Blum, Druke-Noe, Hartung & Köller (2006).

Die Bildungsstandards für die Primarstufe, den Haupt- und mittleren Bildungsgang sind online auf den Webseiten der KMK (<http://www.kmk.org/bildung-schule/qualitaetssicherung-in-schulen/bildungsstandards/ueberblick.html>) oder im Kluwer Verlag verfügbar. Zusätzliche Informationen zur Beschreibung der jeweiligen Kompetenzstufenmodelle sind beim IQB verfügbar ([http://www.iqb.hu-berlin.de/bista?reg=r\\_4](http://www.iqb.hu-berlin.de/bista?reg=r_4)). Als Ergänzung empfiehlt sich die Expertise zur Entwicklung der Standards von Klieme et al. (2003). Ein lesenswertes Plädoyer für die Qualitätssicherung schulischer Arbeit auf der Basis von Bildungsstandards hält Köller (2008).

#### 2.4.4 Verständnis und Diskussionspunkte

1. *Diskutieren Sie, welche Vorteile sich aus einer (bundesweit) zentral vorgenommenen Testentwicklung für Vergleichsarbeiten ergeben. Welche Schwierigkeiten könnten gleichzeitig damit verbunden sein?*
2. *Die Wahl geschlossener Antwortformate, wie etwa Multiple Choice, für die Erfassung von Lese- und Hörverstehen wird auch mit der Validität der Tests begründet. Wie ist dies zu verstehen?*
3. *Im Fach Englisch sind die Testaufgaben an den GER angebunden. Welche Vor- und Nachteile sind mit der Verwendung des GER in diesem Zusammenhang verbunden?*
4. *Vom IQB wird an verschiedenen Stellen betont, dass Testaufgaben keine optimalen Lernaufgaben darstellen. Worin könnten grundlegende Unterschiede bestehen?*

<sup>10</sup>

Der technical report ist als Online-Publikation verfügbar unter: [http://www.iqb.hu-berlin.de/arbberiche/projekte?pg=p\\_7&spg=r\\_8](http://www.iqb.hu-berlin.de/arbberiche/projekte?pg=p_7&spg=r_8)

## 2.5 Praktische Implikationen: Ergebnismutzung aus Vergleichsarbeiten und Unterrichtsentwicklung

Im letzten Kapitel dieses Studienbriefs werden die unmittelbaren Implikationen von Vergleichsarbeiten für die schulische Tätigkeit vorgestellt. Es wird zunächst das Format der Rückmeldungen an einigen Beispielen erläutert. Im Anschluss wird ein Vorschlag zur Umsetzung der aus Vergleichsarbeiten gewonnenen diagnostischen Informationen für pädagogische und didaktische Entwicklungsprozesse (als vielleicht wichtigster Zielsetzung) gemacht. Danach folgt die Beschreibung eines Schulbeispiels, an dem aufgezeigt wird, welchen Einfluss Vergleichsarbeiten auf die Schulpraxis haben können. Abschließend wird die Brücke zum kompetenzorientierten Unterricht geschlagen.

In diesem Kapitel werden folgende Fragen beantwortet:

- *Wie lassen sich die Ergebnismeldungen aus Vergleichsarbeiten interpretieren?*
- *Welche Evaluationsschritte sind im Anschluss an die Ergebnisinterpretation zu durchlaufen, damit Unterrichtsentwicklungsprozesse in Gang gesetzt werden können?*
- *Wie gelangt man von den Ergebnismeldungen der Vergleichsarbeiten zu einem kompetenzorientierten Unterricht?*

### 2.5.1 Interpretation der Rückmeldungen aus Vergleichsarbeiten

Ergebnisinterpretation

Mit der Ergebnisinterpretation aus Vergleichsarbeiten ist das Ziel verbunden, unter der Nutzung kriterialer und sozialer Vergleiche Prozesse der Schul- und Unterrichtsentwicklung in Gang zu setzen. Lehrerinnen und Lehrern der an den Vergleichsarbeiten beteiligten Klassen wird umfangreiches Begleitmaterial zur Verfügung gestellt, welches unter anderem Anregungen zur Nutzung der Testergebnisse und -materialien im Unterricht beinhaltet. Nahezu alle Materialien der Vergleichsarbeiten können in der Regel, wie die Ergebnisse der Schule, über passwortgeschützte Internetportale abgerufen werden, die die Länder zu diesem Zweck eingerichtet haben. Darstellung und Umfang der Rückmeldungen variieren zwischen den Bundesländern. Zumeist umfassen sie aber:

Materialien

- Testaufgaben und Auswertungsmanuale,
- Lösungshäufigkeiten der einzelnen Aufgaben,
- Kommentierungen der Aufgaben und Hinweise für die Weiterarbeit mit den Aufgaben,
- eine Ergebnisdarstellung der eigenen Klassen und Jahrgangsstufen (z.B. als Verteilung von Schülerinnen und Schülern auf Kompetenzstufen),
- kriteriale Beschreibung der Kompetenzanforderungen,
- Angabe von Referenzwerten (beispielsweise Ergebnisse von Schulen, welche unter vergleichbaren Rahmenbedingungen arbeiten) sowie (statistische) Interpretationshilfen für einen sozialen Vergleich.

Diagnostisches Potential von Vergleichsarbeiten

Das diagnostische Potential der Testaufgaben liegt darin, differenziert Stärken und Schwächen der Schülerschaft in Bezug auf einzelne Inhalte zu untersuchen (vgl. Möller, Pallack & Fleischer, 2007; Kliemann, 2010). Dennoch soll im Folgenden die kriteriale und soziale Interpretation im Hinblick auf *Kompetenzausprägungen* im Zentrum der Darstellung von Ergebnismeldungen stehen. Um die Darstellung nicht zu abstrakt werden zu lassen, wird wiederum auf konkrete Beispiele der Rückmeldungen aus den vergangenen Jahren zurückgegriffen.

#### Beispiel:

Abbildung 9 (S. 29) zeigt eine Schulrückmeldung des Kompetenzbereichs Leseverstehen im Unterrichtsfach Deutsch, Kompetenzbereich Leseverstehen, der Lernstandserhebungen in Nordrhein-Westfalen im Jahre 2011. Die Rückmeldung beinhaltet für eine ausgewählte Klasse (oben), die Jahrgangsstufe (Mitte) und die Parallelklassen (unten) der Beispielschule die prozentuale Verteilung der Schülerinnen und Schüler auf fünf Kompetenzstufen zuzüglich einer Kategorie „k.h.N.“. Liegen für eine Schülerin/einen Schüler keine oder unvollständige Daten vor, ist eine Zuordnung zu den beschriebenen Kompetenzniveaus nicht möglich; dies wird als „kein hinreichender Nachweis für das Erreichen eines Kompetenzniveaus“ bezeichnet, d. h. in diesem Bereich wurden in der Regel nur vereinzelt Aufgaben gelöst. Die insgesamt geringe Anzahl dieser gelösten Aufgaben ermöglicht es jedoch nicht, die Schülerin bzw. den Schüler mit hinreichender Sicherheit einem Kompetenzniveau zuzuordnen. Der rechte Rand der Abbildung zeigt Erläuterungen zu den Kompetenzstufen. Lehrerinnen und Lehrer werden hier zu den Kompetenzanforderungen der jeweiligen Stufe weitergeleitet.

Der Abbildung ist zunächst im Sinne eines kriterialen Vergleichs zu entnehmen, dass der überwiegende Teil der ausgewählten Klasse (73 %) die Anforderungen der Kompetenzstufen 3, 4 und 5 nicht erfüllt. Lediglich 17% Schülerinnen und Schüler erreichen die Stufe 3, 9 % der Schülerinnen und Schüler erzielen Leistungen auf Stufe 4. Ein ähnliches Ergebnis findet sich auch für die gesamte Jahrgangsstufe (Mitte), in der 78% der Schülerinnen und Schüler nicht die Kompetenzstufe 3 erreichen. Für eine möglichst objektive Einschätzung dieser Ergebnisse ist es bedeutsam, mit welcher Referenzgruppe sich die Schule vergleicht. Die Schule ist auf Basis des in Kapitel 2.2.2.1 vorgestellten Verfahrens der Referenzgruppe einem sogenannten Standorttyp, dem Standorttyp 5 mit den ungünstigsten Eingangs Voraussetzungen der Schülerschaft, zugeordnet worden. Diese Referenzgruppe ermöglicht einen Vergleich mit anderen Schulen, an denen unter ähnlichen Rahmenbedingungen gearbeitet wird.

## Deutsch Leseverstehen

23 von 23 Schülerinnen und Schülern der Klasse D-8a haben teilgenommen.

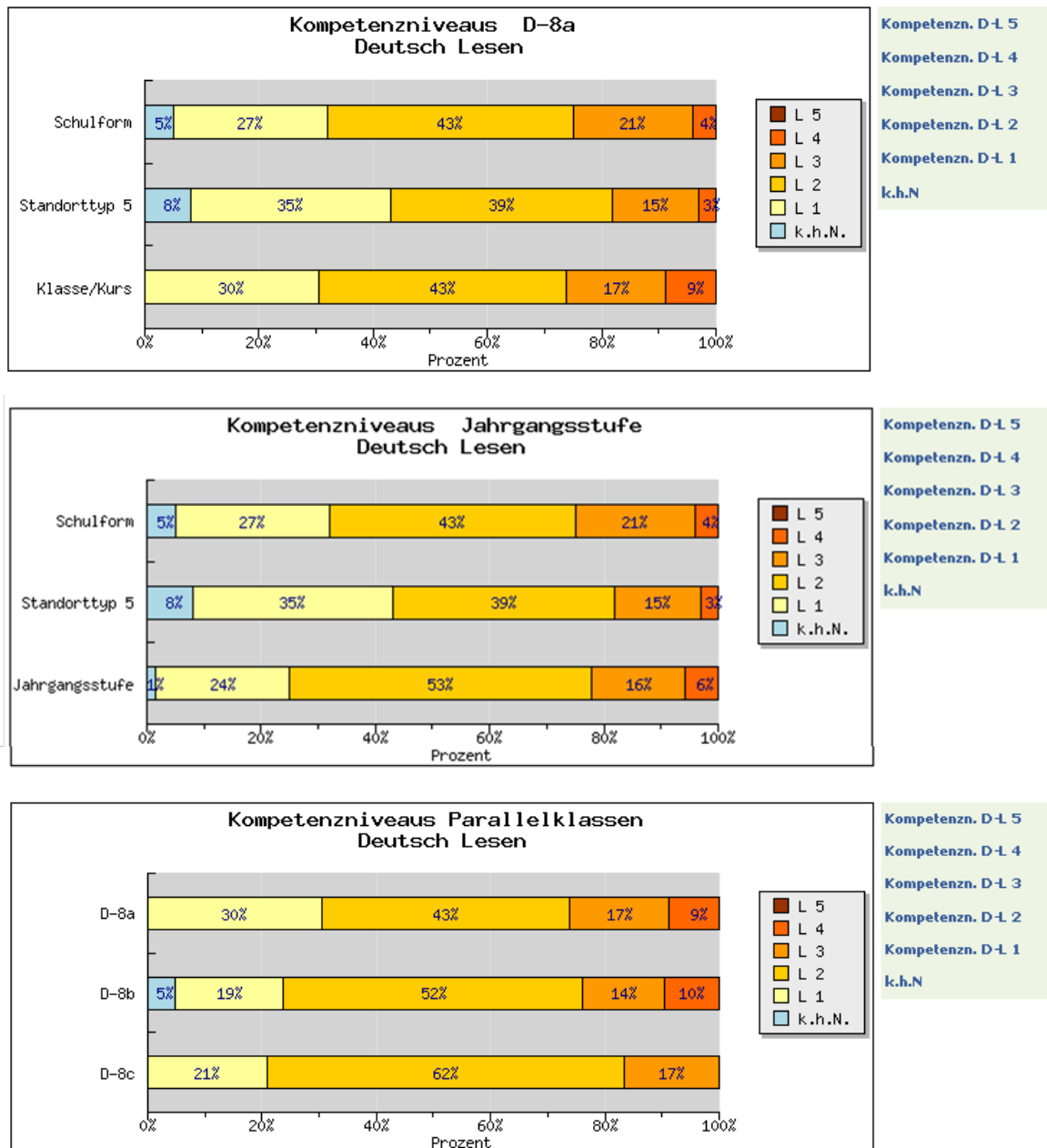


Abbildung 9: Ergebnissrückmeldung Lernstand 8 in Nordrhein-Westfalen des Jahres 2011 (Beispiel für eine Klasse im Fach Deutsch, Kompetenzbereich Leseverstehen)

Die Berücksichtigung der Referenzverteilungen von Standorttyp 5 im Sinne eines sozialen Vergleichs relativiert das Abschneiden der Klasse und Jahrgangsstufe, denn in diesem Standorttyp verfehlen insgesamt sogar 82% das Kompetenzniveau 3. Um zu verhindern, dass marginale und eher zufällige Unterschiede zwischen der eigenen und den Referenzverteilungen interpretiert werden, erhalten Lehrkräfte zu den Abbildungen jeweils kurze Hinweistexte.

Der Vergleich mit den Parallelklassen (unten) liefert viele Hinweise für die innerschulische Diskussion, da innerhalb der Schule detaillierte Informationen zu dem in den Parallelklassen realisierten unterrichtlichen Vorgehen verfügbar sind. Bei der vorgestellten Schule fällt beispielsweise auf, dass die Klasse 8c homogener abschneidet als ihre Parallelklassen, deren Ergebnisse stärker streuen.

In Kapitel 2.2 war bereits darauf eingegangen worden, dass bei den Kompetenztests in Thüringen ein anderer Weg der Ergebnissrückmeldung gewählt worden ist. Die Klassen erhielten bei den Kompetenztests die durchschnittlich erreichte absolute und prozentuale Punktzahl zurückgemeldet. Diese Punktzahlen konnten mit einem auf Basis schulleistungsrelevanter Hintergrundmerkmale korrigierten Landesmittelwert (siehe Kapitel 2.2) verglichen werden. Ab-

bildung 10 (S. 30) zeigt eine Ergebnismeldung des thüringischen Kompetenztests in der dritten Jahrgangsstufe aus dem Jahr 2009.

Die Abbildung führt auf der linken Seite in Form eines Balkendiagrammes die durchschnittliche prozentuale Punktzahl im Test für die Beispielklasse (rot) und den korrigierten Landesmittelwert (gelb) auf. Dieser korrigierte Landesmittelwert ist aufgrund sozioökonomischer Hintergrundmerkmale der Schülerschaft in der Beispielklasse 3Z berechnet worden und ermöglicht einen „fairen“ sozialen Vergleich der Schulleistungen in dieser Klasse (vgl. Kapitel 2.2). Auf der rechten Seite erhält die Lehrkraft zusätzlich einige deskriptive Statistiken<sup>11</sup> der erzielten Ergebnisse. Sie kann sich darüber beispielsweise einen Eindruck von der Leistungsstreuung in ihrer Klasse (minimal erzielte Punktzahl/ maximal erzielte Punktzahl) machen und erfährt die maximal erreichbaren Punkte sowie die absolute und relative Lösungsquote der eigenen Klasse und des korrigierten Landesmittelwerts.

Dem Balkendiagramm ist zu entnehmen, dass in der Beispielklasse fast 80 % der Aufgaben gelöst wurden und sie damit im Vergleich mit dem (korrigierten) Landesdurchschnitt (72 %) besser abscheidet. Den deskriptiven Statistiken lässt sich des Weiteren entnehmen, dass der Unterschied zwischen der Klasse und dem (korrigierten) Landesdurchschnitt bei 2,5 Punkten liegt. Die Streuung in der Beispielklasse umfasst 24 Punkte (59 %) und erscheint damit auf den ersten Blick recht groß. So erzielt die schwächste Schülerin bzw. der schwächste Schüler der Klasse mit 16 gelösten Items gut 40 % der möglichen Punkte, die beste Schülerin oder der beste Schüler verfehlt mit 40 Punkten (98 %) nur knapp die Höchstpunktzahl.

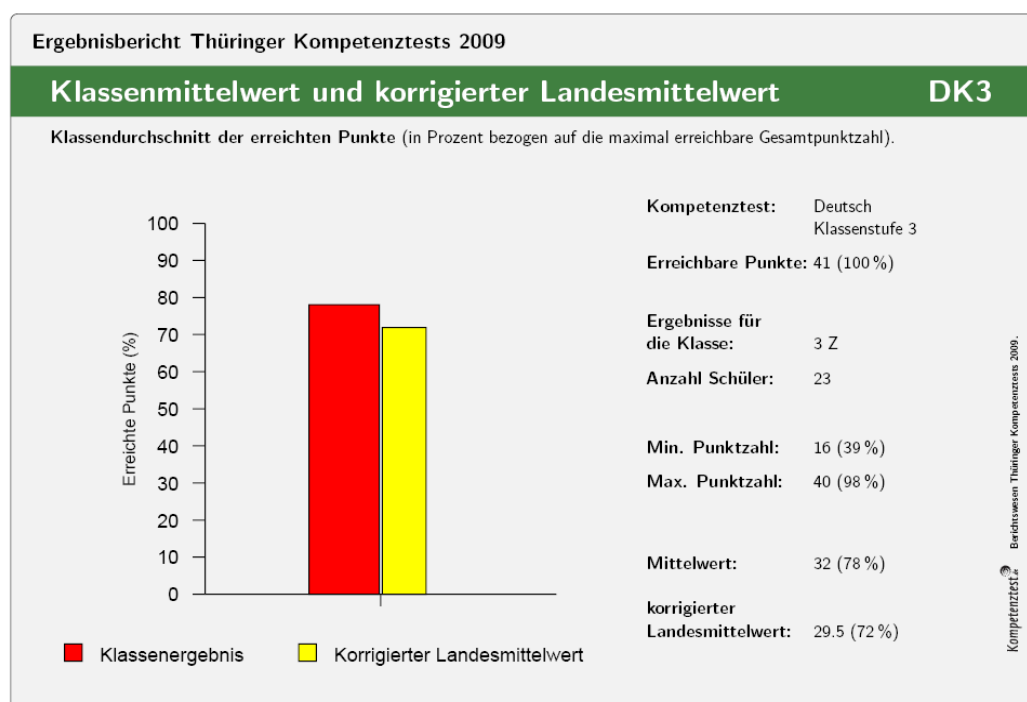


Abbildung 10: Ergebnismeldung der „Kompetenztest“ in Thüringen 2009 (Beispiel für eine Klasse im Fach Deutsch)

Weitere Informationen über die Leistungsstreuung innerhalb der Klasse sind der Abbildung 11 (S. 31) zu entnehmen, die sich ebenfalls auf die Beispielklasse 3Z bezieht. Die Abbildung zeigt einen so genannten Boxplot der Leistungsstreuung in der Beispielklasse (hervorgehoben oben) und dem Landesmittelwert (unten). Das mittlere graue Segment des Boxplots umfasst die mittleren 50 % der Schülerinnen und Schüler in der Klasse bzw. des Landes, wobei durch einen senkrechten Strich der Median angegeben ist. Der Median beschreibt jenen Testwert, unter und über dem 50 % der getesteten Schülerinnen und Schüler liegen; er halbiert also die Verteilung. Links des grauen Segmentes und der Markierung „25 %“ liegt das Viertel der leistungsschwächsten Schülerinnen und Schüler. Rechts des grauen Segmentes und der Markierung „75 %“ liegt das Viertel der leistungsstärksten Schülerinnen und Schüler. Das Ende des roten bzw. blauen Bereiches markiert den Beginn der extremsten 10 % der Klasse bzw. des Landes, die dahinter liegenden „Whiskers“ (Schnurrbarthaare) geben in der Verteilung der Klasse die jeweils leistungsschwächste (Minimum) bzw. leistungsstärkste (Maximum) Person an. Die Abbildung verdeutlicht, dass im Landesdurchschnitt die Streuung der Testergebnisse in der unteren Leistungshälfte (rechts des Median) größer ist als in der oberen Leistungshälfte. Insbesondere das leistungsstärkste Quartil des Landes ist in seinen Leistungen vergleichsweise homogen, was allerdings nicht überrascht, da die Lösungsquote des Tests relativ hoch ist. Dieses Ergebnis spiegelt sich auch beim Blick auf den Boxplot der Beispielklasse wider. Revidiert werden muss hingegen der erste Eindruck von der Leistungsstreuung in Klasse 3Z, denn der Abstand zwischen den stärksten und schwächsten 10 % der Beispielklasse ist geringer als im Landesdurchschnitt. Der Lehrkraft in der Beispielklasse ist es also offenbar gelungen, die

<sup>11</sup> Deskriptive Statistiken fassen die Ergebnisse der Klasse anhand geeigneter statistischer Maße (z.B. Mittelwert und Standardabweichung) zusammen.



Leistungsstärke in der Klasse zu homogenisieren. Es kann an diesem Beispiel verdeutlicht werden, wie bedeutsam die Berücksichtigung von Referenzwerten ist: Wird der soziale Vergleich ignoriert, können Fehlinterpretationen entstehen. Der Abbildung ist zudem zu entnehmen, dass die Beispielklasse insbesondere im unteren Leistungsbereich (rotes Segment des Boxplots) homogener abschneidet als der Landesdurchschnitt. Insbesondere recht schwache Leistungen sind also in der Beispielklasse selten. Im Median unterscheiden sich beide Boxplots dagegen kaum. Dies bestätigt den Eindruck, dass das bessere durchschnittliche Abschneiden der Beispielklasse im Wesentlichen auf die geringere Streuung im unteren Leistungsbereich zurückzuführen ist.

Die Beispiele verdeutlichen, dass für eine verständnisvolle Nutzung der Ergebnismeldungen eine Auseinandersetzung mit dem jeweiligen Rückmeldeformat notwendig ist. So lassen sich aus den Abbildungen und den zusätzlichen Informationen der beiden Beispiele unterschiedliche Erkenntnisse über das Abschneiden der eigenen Klasse (und ggf. der eigenen Jahrgangsstufe) ziehen. Statistische Grundkenntnisse werden dabei vorausgesetzt. Lehrkräften, denen diese fehlen, kann als Literatur z. B. Eikenbusch und Leuders (2004) oder Lienert und von Eye (1998) empfohlen werden.

Die Rückmeldeformate der Bundesländer variieren, sodass die Formate in einigen Ländern von den hier dargestellten Beispielen abweichen können. In der Regel stellen die Länder aber auf ihren Webseiten Beispiele für die Ergebnisinterpretation der landespezifischen Formate bereit.

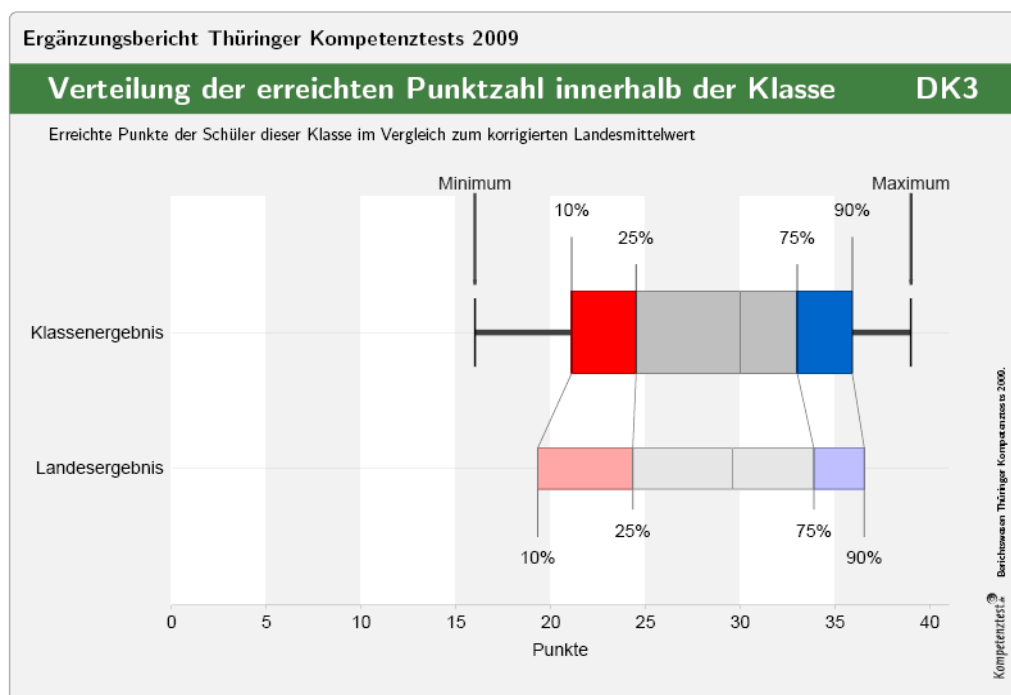


Abbildung 11: Ergänzende Informationen zur Ergebnismeldung der „Kompetenztests“ in Thüringen 2009 (Beispiel für eine Klasse im Fach Deutsch)

### 2.5.2 Von der Ergebnisinterpretation zur schulischen Entwicklung

Wie kann die zentrale Zielsetzung der Unterrichtsentwicklung auf Basis der Ergebnisinterpretation aus Vergleichsarbeiten realisiert werden? Im Folgenden werden vier zentrale Schritte zur Nutzung der Ergebnismeldungen für eine schulinterne Evaluation vorgestellt. Sie sind grob am Zyklusmodell von der Evaluation zur Innovation nach Helmke und Hosenfeld (2004) orientiert, werden aber für die entsprechende Zielsetzung hier konkretisiert<sup>12</sup>.

#### 1) Ergebnisinterpretation anhand kriterialer und sozialer Bewertungsmaßstäbe:

Eine zutreffende Interpretation ist Voraussetzung für die verständnisvolle Ergebnismeldung. Die kriteriale Interpretation einer Ergebnismeldung beinhaltet beispielsweise folgende Fragen (vgl. auch Peek & Dobbelstein, 2006):

*Wie viele meiner Schüler erreichen welche Kompetenzerwartungen?*

*Lassen sich anhand der Fehlermuster typische Kompetenzlücken identifizieren?*

Relevant wäre z.B., ob Fehler systematisch in bestimmten **Inhaltsbereichen** oder bei bestimmten **Aufgabentypen** auftreten. Vor dem Hintergrund sozialer Vergleiche kann eine Lehrkraft untersuchen:

*Wie schneidet meine Klasse innerhalb der Jahrgangsstufe ab? Bestehen Leistungsunterschiede zwischen meiner Klasse und den Parallelklassen? Sind Stärken und Schwächen in bestimmten Bereichen identifizierbar (differenzielle Leistungsunterschiede)? Wie schneidet meine Klasse im Vergleich mit Klassen anderer Schulen mit ähnlicher Schülerzusammen-*

Ergebnis-  
interpretation

kriteriale  
Interpretation

soziale  
Vergleiche

<sup>12</sup> Neben dem hier beschriebenen Rezeptionsverlauf beinhaltet das Modell individuelle und externe Rahmenbedingungen, die hier nicht angeführt werden.

setzung ab? Finden sich Hinweise, dass meine Lerngruppe entsprechend dieser Referenz ihre Möglichkeiten nicht ausschöpft?

Zudem bietet es sich an, die Ergebnisse zu Zeugnisnoten, Klassenarbeiten und Parallelarbeiten ins Verhältnis zu setzen. Dabei sollte das Gesamtbild der Klasse betrachtet werden und bzgl. systematischer Abweichungen geprüft werden.

## 2) Reflexion und Ursachenanalyse in den Fachkonferenzen:

Individuelle  
Reflexion und  
Diskussion der  
Ergebnisse

Vergleichsarbeiten sollen Lehrkräfte in die Lage versetzen, in den Fachkonferenzen über die Ursachen für die festgestellten Ergebnisse in Diskussion zu kommen und Anhaltspunkte für mögliche Interventionen herauszuarbeiten. Es sollten insbesondere berücksichtigt werden:

- Besonderheiten der Schülerschaft
- inhaltliche Überschneidungen mit anderen Unterrichtsfächern, z.B. durch bilingualen Unterricht oder Überschneidungen zwischen Mathematik- und Informatikunterricht

## 3) Folgerungen aus den Ergebnissen und Interventionen:

Folgerungen  
aus den  
Ergebnissen /  
Interventionen

Im Sinne des Wechsels von einer Input- zu einer Outputorientierung im Bildungswesen liegt es im Verantwortungsbereich der Schule, unter Berücksichtigung der eigenen schulischen Besonderheiten selbstständig geeignete Konsequenzen aus den Ergebnismeldungen zu entwickeln. Es sollen hier nur grob einige Ansatzpunkte beschrieben werden:

- Gemeinsame Unterrichtsvorbereitung oder Hospitationen in Parallelklassen können den eigenen Referenzrahmen für die zu erwartenden Leistungen von Schülerinnen und Schülern erweitern und die eigene diagnostische (Bewertungs-)Kompetenz stärken. Unterschiedliche Unterrichtsmaterialien und Leistungsanforderungen der Fachlehrerinnen und -lehrer können zudem einander gegenübergestellt werden.
- Die Passung zwischen schulinternem Curriculum und den in den Bildungsstandards skizzierten Kompetenzanforderungen sollte untersucht werden. Didaktische Trainer können sinnvolle Unterstützung bieten, um neue Inhalte in das schulinterne Curriculum zu integrieren.
- Individuelle Fördermaßnahmen, insbesondere für die Gruppe der leistungsschwachen Schülerinnen und Schüler, gewinnen oftmals an Bedeutung. In diesem Fall ist zu prüfen, ob sich die Ergebnisse der Vergleichsarbeiten in ein konsistentes Bild über diese Schülerinnen und Schüler einfügen lassen, was eine systematische Dokumentation der Leistungsentwicklung voraussetzt. Im Anschluss können geeignete Fördermaßnahmen (Förderstunden, Binnendifferenzierung, etc.) in Erwägung gezogen werden.

## 4) Schulinterne Evaluation der Interventionsmaßnahmen:

Evaluation von  
Interventions-  
maßnahmen

Hat eine Intervention überhaupt Wirkung gezeigt und wie nachhaltig ist die Wirkung? Um dies beurteilen zu können, bietet es sich zum Beispiel an, zu einem späteren Zeitpunkt dieselben oder parallele Item-Stichproben der Vergleichsarbeiten erneut zu verwenden. Es kann so anhand repräsentativer Lösungsquoten erfasst werden, wie groß der Lernfortschritt von Schülerinnen und Schülern in einem bestimmten Zeitintervall ist und ob sich Profile von Kompetenzstärken und -schwächen verändert haben. Lehrerinnen und Lehrern werden künftig verstärkt Materialien zur Lernstandsdiagnose bereitgestellt. Das IQB plant, den Schulen zum Zweck der Selbstevaluation den Abruf von identischen oder parallelen Testaufgaben der Vergleichsarbeiten in der achten Jahrgangsstufe aus einer Itemdatenbank zu ermöglichen. Ein ähnliches Vorhaben verfolgen in der Mathematik auch die Projekte "SINUS-Transfer" (Sekundarstufe I; vgl. auch Studienbrief 3) und „SINUS Transfer Grundschule“, welche ebenfalls Materialien zur regelmäßigen Lernstandsdiagnose zur Verfügung stellen. Lehrerinnen und Lehrer haben so die Möglichkeit, zu einem ihnen günstig erscheinenden Zeitpunkt die Entwicklung des Kompetenzerwerbs ihrer Schülerinnen und Schüler festzustellen.

Mit den hier vorgestellten vier Schritten der internen Evaluation ist der schulische Entwicklungsprozess natürlich nicht abgeschlossen. Die Evaluationsbefunde gehen vielmehr in die Interpretation weiterer diagnostischer Informationen ein, sodass sich ein Zyklus regelmäßiger Evaluationsmaßnahmen ergibt. Abbildung 12 (S. 33) fasst die vier vorgestellten Schritte noch einmal zusammen.

### 2.5.3 Ein Beispiel für die Ergebnisnutzung in den Fachkonferenzen

Fachkonferenzen dürfen in den Schulen als ein zentrales Gremium der Qualitätsentwicklung betrachtet werden. In ihrem Aufgabenbereich liegen (i.d.R.) zahlreiche didaktische und methodische Arbeiten, u.a.:

- die Ausarbeitung des schulinternen Curriculums und gemeinsamer Unterrichtsreihen
- die Evaluation geeigneter Unterrichtsmaterialien
- die Organisation von Hospitationen und Weiterbildungen.

Wie bereits dargelegt (vgl. den vorangegangenen Abschnitt), sind Fachkonferenzen mit diesem Verantwortungsbereich auch geeignete Adressaten der Rückmeldungen aus Vergleichsarbeiten. Zur Verdeutlichung, wie die Ergebnisse der Vergleichsarbeiten in den Fachkonferenzen diagnostisch genutzt werden können, wird im Folgenden die Arbeit

einer Fachkonferenz Deutsch im Anschluss an eine Ergebnissrückmeldung zum Kompetenzbereich „Leseverstehen“ als Beispiel dargestellt.

In der Auseinandersetzung mit den Ergebnissen hatte die Fachkonferenz auf Basis von Vorarbeiten der Jahrgangsstufenlehrerinnen und -lehrer beschlossen, folgende Punkte zu thematisieren:

- Allgemein: Welche Ergebnisse bedürfen einer besonderen Beachtung?
- Wie gestaltet sich der Parallelklassenvergleich? Lassen sich aus den Ergebnissen Informationen gewinnen, welche in den schulinternen Vergleichen (bspw. Parallelarbeiten) nicht offensichtlich geworden sind?
- Wie gestaltet sich der Vergleich der Jahrgangsstufe mit Referenzschulen?
- Ergeben sich systematische Unterschiede zwischen Schülergruppen, bspw. zwischen Mädchen und Jungen?
- Welche Konsequenzen ergeben sich aus den Ergebnissen für die vorausgegangenen bzw. die nachfolgenden Jahrgangsstufen? Wie lässt sich in diesen Jahrgangsstufen kompetenzorientierter Unterricht erfolgreich implementieren bzw. weiterentwickeln?

#### REZEPTION UND INTERPRETATION DER ERGEBNISRÜCKMELDUNG

##### Kriteriale Bewertungsmaßstäbe:

- Wieviele Schüler erreichen welche Kompetenzerwartungen?
- Lassen sich anhand der Fehlermuster typische Kompetenzlücken identifizieren?

##### Soziale Bewertungsmaßstäbe:

- Wie schneidet meine Klasse innerhalb der Jahrgangsstufe ab?
- Bestehen deutliche Diskrepanzen in den erreichten Kompetenzerwartungen zwischen meiner Klasse und den Parallelklassen?

#### REFLEKTION DER ERGEBNISSE UND URSACHEN, INDIVIDUELL UND IN DEN FACHGRUPPEN

Ergebnissrückmeldungen geben Anhaltspunkte für mögliche Interventionen.

Erklärungssuche:

- unpassende fachliche Vorbereitung
- Abweichungen zwischen schulinternem Curriculum und Kernlehrplänen

#### FOLGERUNGEN UND INTERVENTIONEN

Förderung des Kompetenzerwerbs von Schülerinnen und Schülern:

- Gemeinsame Unterrichtsvorbereitung oder Hospitationen in Parallelklassen
- Überprüfung des schulinternen Curriculums (inhaltliche Schwerpunktsetzung)
- Individuelle Fördermaßnahmen für leistungsschwache Schüler/innen

#### (SCHULINTERNE) EVALUATION

Evaluation der umgesetzten Interventionsmaßnahmen:

- Wie erfolgreich waren die Interventionen?
- Ist eine Änderung des Profils von Stärken und Schwächen erkennbar?
- Lassen sich Subgruppen identifizieren, die mehr oder weniger von den Interventionen profitiert haben?
- Wie nachhaltig ist die Wirkung der Interventionen?

Abbildung 12: Schema für das Vorgehen von der Interpretation der Ergebnissrückmeldung bis zur Ergebnisnutzung für die schulinterne Evaluation

Zwei der hier aufgeführten Diskussionspunkte seien exemplarisch herausgegriffen, um konkret darzustellen, welche Konsequenzen die Fachkonferenz aus den Ergebnissen gezogen hat:

Der Parallelklassenvergleich der Ergebnisse zeigte deutliche Unterschiede zwischen den Klassen. So schnitten zwei der vier Parallelklassen signifikant besser ab als entsprechende Referenzgruppen, wohingegen die anderen beiden Klassen der Gesamtschule im Durchschnitt blieben. Die Fachkonferenz verglich die erzielten Ergebnisse mit früheren Leistungsvergleichen der Klassen, welchen derartige Leistungsunterschiede jedoch nicht zu entnehmen waren. Diskutiert wurden Unterschiede zwischen den Klassen in der Zusammensetzung der Schülerschaft (Lernmotivation der Klasse, Anteil der Deutsch-Muttersprachler, Mädchenanteil; vgl. den nachfolgenden Absatz). Die Mitglieder zeigten sich jedoch skeptisch, ob Unterschiede in der Schülerzusammensetzung die Ergebnisse allein erklären könnten. Zur Sprache kam im Anschluss auch die Frage, inwiefern eine unterschiedliche Bewertungsstrenge der Fachlehrerinnen und -lehrer eine Rolle spielen könnte. Als weitreichende Konsequenz schlug die Fachkonferenz der Schulleitung die Einführung regelmäßiger Parallelarbeiten vor, um einen besseren schulinternen Vergleich zu ermöglichen und innerhalb der Fachkonferenz einen noch intensiveren Austausch über Unterrichtsinhalte, Aufgaben und Bewertungs-



kriterien anzuregen. Nach anfänglicher Skepsis werden diese in den Kernfächern inzwischen in allen Jahrgangsstufen der Sekundarstufe I einmal pro Schuljahr durchgeführt und stehen im Kollegium inzwischen als hilfreiches Instrument des sozialen Vergleichs außer Zweifel. Der Nutzen von Vergleichsarbeiten für einen darüber hinaus gehenden externen Vergleich mit anderen Schulen ist ebenfalls richtig wahrgenommen worden.

Ein weiteres interessantes Ergebnis der Vergleichsarbeiten beinhaltete Geschlechtsunterschiede in der Lesekompetenz: Den Ergebnismeldungen der Gesamtschule war deutlich zu entnehmen, dass Jungen im Durchschnitt schwächer abschnitten als Mädchen. Die Fachkonferenz bezog in die Diskussion der Ergebnisse der Vergleichsarbeiten auch Ergebnisse eines standardisiert durchgeführten Sprachstandtests zu Beginn der fünften Jahrgangsstufe ein, die im Lesen ebenfalls einen deutlichen Geschlechtsunterschied zugunsten der Mädchen feststellen ließen. Offenbar wurden schwächere Leseleistungen der Jungen also aus der Grundschule „verschleppt“. Ausgehend von der Feststellung, dass schwächere Leseleistungen nicht zuletzt auch aus einem geringeren Leseinteresse und einer niedrigeren Lesefrequenz erklärt werden können, wurde die pragmatische Entscheidung getroffen, dass bis zu drei Jungen aus jeder Klasse in der 5. und 6. Jahrgangsstufe eine zusätzliche Stunde Leseunterricht (Viellese-Verfahren) erhalten sollten. Die Schüler wählen sich dazu jeweils eigenständig Bücher und trainieren, begleitet von einem männlichen Lehrer, Lesefertigkeiten und Lesestrategien. Die Fachkonferenz entschied, diese Maßnahme zu evaluieren, indem die Leseleistungen von Jungen und Mädchen nun regelmäßig in allen Jahrgangsstufen verglichen werden. Der bei den Vergleichsarbeiten NRW in regelmäßigen Zyklen wiederkehrende Kompetenzbereich „Leseverstehen“ eröffnet der Fachkonferenz zusätzlich die Möglichkeit, die verwendeten Leseaufgaben zu reflektieren und Referenzwerte für die Leseleistungen von Jungen und Mädchen zu erhalten.

Im Anschluss widmete sich die Fachkonferenz der Frage, wie Methoden des kompetenzorientierten Unterrichts stärker zu implementieren seien und setzte sich dazu mit den die Ergebnismeldung ergänzenden Materialien zur Weiterarbeit auseinander. Die Implementation kompetenzorientierten Unterrichts wird, veranschaulicht am Kompetenzbereich Leseverstehen, separat im folgenden Abschnitt behandelt.

Die Fachkonferenz beschloss ihre Arbeit mit der Ausarbeitung eines Ergebnisberichts an die Schulaufsicht. In ihm werden neben den Ergebnissen auch die spezifische Situation der Schule und die geplanten Interventionsmaßnahmen dargelegt.

#### 2.5.4 Kompetenzorientierter Unterricht

Grundgedanke kompetenzorientierten Unterrichts

Vergleichsarbeiten ebnen den Weg für kompetenzorientierten Unterricht. Unter kompetenzorientiertem Unterricht wird hier nach Ziener (2006) verstanden, "Stoffe, Inhalte oder Themen im Unterricht so zu bearbeiten, dass dabei Kompetenzen, wie sie in den Bildungsstandards formuliert sind, angebahnt, eingeübt oder erworben werden können". Nach den Vorstellungen der KMK (KMK & for.mat, ohne Jahr) verläuft die Planung kompetenzorientierten Unterrichts in drei Schritten:

1. Aneignung von Kenntnissen über die Inhalte der Bildungsstandards und Verständnis dieser Inhalte: Dies bezieht sich nicht allein auf die Kompetenzbeschreibungen, sondern beinhaltet auch die in den Standards aufgeführten Methoden und Arbeitstechniken (vgl. Kapitel 2.4).
2. Untersuchung der didaktischen Ziele dieser Kompetenzbeschreibungen: Die KMK spricht hier von der „Kompetenzexegese“, deren Inhalt nicht zuletzt die gegenstandsbezogene Konkretisierung der Kompetenzbeschreibungen ist.
3. Ableitung von Aufgabenstellungen und Erstellung einer Sequenzplanung durch die strukturierte Zusammenführung dieser Aufgabenstellungen: Die Aufgabenbeispiele der Vergleichsarbeiten geben Lehrkräften bereits einen Eindruck davon, wie kompetenzorientierte (Test-)Aufgaben aussehen können. Es liegt in der Verantwortung der Fachkonferenz, diese zu kompetenzorientierten Lernaufgaben weiterzuentwickeln und in geeignete Unterrichtsreihen einzubinden. Sie diskutiert, welche Veränderungen der Materialien (Kürzungen oder Erweiterungen) notwendig sind, und mit welchen Konsequenzen in Bezug auf Diagnose und Differenzierung der Einsatz der Materialien verbunden ist. Häufig bietet es sich an, innerhalb der Fachkonferenz Tandems von jeweils zwei Lehrkräften zu bilden, welche jeweils bestimmte Teile der Materialien und Aufgaben bzgl. ihrer Potentiale untersuchen und Erfahrungen im Umgang mit den Materialien dokumentieren.

Kompetenzzuwächse im Unterricht

Um Kompetenzzuwächse im Unterricht zu erzielen, wird von der KMK die Verknüpfung zweier Elemente vorgeschlagen (KMK & for.mat, ohne Jahr): zum einen die regelmäßige diagnostische Erfassung der Lernstände der Schülerinnen und Schüler, etwa im Rahmen von Vergleichsarbeiten. Zum anderen wird davon ausgegangen, dass Kompetenzzuwächse durch eine systematische Variation des Schwierigkeitsgrades des Unterrichtsgegenstandes bzw. der im Unterricht eingesetzten Aufgaben erzielt werden. Im KMK-Projekt for.mat (<http://www.kmk-format.de>) sind Analyseschemata entwickelt worden, welche es möglich machen, schwierigkeitsbestimmende Merkmale von Texten und Leseaufgaben zu identifizieren. Abbildung 13 (S. 35) gibt hierzu ein Beispiel für Lernaufgaben des Kompetenzbereichs Leseverstehen. Alternativ lässt sich die Schwierigkeit des Unterrichtsgegenstandes variieren, im Bereich Leseverstehen also beispielsweise die Textschwierigkeit. Beide Methoden stellen eine nicht zu unterschätzenden Herausforderung an die diagnostischen Fähigkeiten der Lehrerinnen und Lehrer dar, denn die Schwierigkeit von Unter-

richtsgegenstand und Aufgaben muss von ihnen richtig eingeschätzt werden. Entsprechende Hilfsmittel sind allerdings inzwischen entwickelt worden (vgl. ebenfalls die zuvor bereits beschriebene Webseite von for.mat).

Im Folgenden werden zwei Beispiele zur Weiterarbeit im Unterricht vorgestellt.

### Beispiel 1: Möglichkeiten zur Förderung der Lesekompetenz im Deutschunterricht am Beispiel der Hierarchie-Ebenen „lokale Kohärenzbildung“ und „Bildung von Superstrukturen“

Richter und Christmann (2002) unterscheiden in ihrer Ausgestaltung eines kognitionstheoretischen Modells des Lesens verschiedene Hierarchie- oder Anforderungs-Ebenen (vgl. Tabelle 9, S. 35), welche jeweils zum Ausgangspunkt für die Förderung der Lesekompetenz werden können (IQB, 2009; vgl. auch Rosebrock & Nix, 2008). Exemplarisch seien hier die Ebenen „lokale Kohärenzbildung“ als eine hierarchieniedrige und „Bildung von Superstrukturen“ als eine hierarchiehohe Prozessebene herausgegriffen. Unter „lokaler Kohärenzbildung“ werden Prozesse verstanden, bei denen die Schülerinnen und Schülern aus einzelnen Satzfolgen Sinnzusammenhänge bilden. Dies geschieht unter Berücksichtigung semantischer und syntaktischer Gesichtspunkte des Textes. Kontextspezifisches Vorwissen erleichtert es dabei Inferenzen zu bilden, also Leerstellen im Text sinnvoll zu ergänzen. Auf dieser hierarchieniedrigen Ebene können Viellese-Verfahren ein Baustein zur Förderung der Leseroutine und Lesemotivation sein (vgl. Rodebrock & Nix, 2008): Im Unterricht werden verbindliche Termine festgelegt, zu denen die Schülerinnen und Schülern individuell ausgewählte und im Unterricht nicht behandelte Texte der Kinder- und Jugendliteratur lesen. Schwache Leserinnen und Leser, die über keine Lesemotivation verfügen, werden durch diese Verpflichtung an das Lesen herangeführt. Geübtere Leserinnen und Leser können ohne Leistungsdruck eigenen Lesewünschen nachgehen, woraus zusätzliche Lesemotivation erwächst.

Förderung  
der Lese-  
kompetenz  
im Deutsch-  
unterricht

schwierigkeitsbestimmendes Merkmal		Ausprägungsgrad			
		sehr gering	gering	eher hoch	hoch
Komplexität der Aufgabenstellung	• Integrationsgrad (Grad der Menge und Dichte notwendiger Schlussfolgerungen)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Implizitheitsgrad der Operatoren / der Arbeitsanweisungen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Komplexität und Anforderungshöhe des Bezugsgegenstandes (Text/Textensemble/Problemstellung)	• Diversität des Bezugsgegenstands	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Anspruchsniveau des Bezugsgegenstands	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Anforderungen an Schülervoraussetzungen bzgl.	• Weltwissen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Fachwissen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Anforderungen an die sprachliche Darstellung der Produkte bzgl.	• Sprachliche Gestaltung (Lexik, Syntax)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Vielfalt der erwarteten Produktmerkmale	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Umfang und Komplexität der notwendigen Reflexion und Bewertung	• Differenziertheit des Urteils	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Verknüpfungsgrad der Bewertungsaspekte	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	• Grad des Rückbezugs auf Bewertungsmaßstäbe	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Abbildung 13: Analyseschema für schwierigkeitsbestimmende Merkmale von Lernaufgaben des Kompetenzbereichs „Leseverstehen“ im Fach Deutsch

Um auch eine Verbesserung der Lesekompetenz zu erzielen, ist das Viellese allein möglicherweise nicht ausreichend (vgl. Rosebrock & Nix, 2008). Angesichts der Bedeutung des Vorwissens sollte daher neben dem textsortenbezogenen Wissen auch grammatisches und lexikalisches Wissen im Unterricht gezielt angesprochen werden. Eine Möglichkeit dazu bietet etwa die Wortschatzarbeit in einer Unterrichtssequenz mit idiomatischen Wendungen (vgl. IQB, 2009):

1. Zu Beginn der Unterrichtssequenz wird beispielhaft ein ausgewähltes Sprichwort behandelt.
2. Davon ausgehend diskutiert die Klasse allgemein die Bedeutung von Sprichwörtern.
3. Im Sinne der Wortschatzerweiterung analysieren die Schülerinnen und Schüler eine bestimmte Anzahl von Sprichwörtern in Gruppenarbeit. Die Gruppen bereiten eine visuelle Darstellung ihrer Ergebnisse vor.
4. Schließlich präsentieren die Gruppen ihre Arbeit der Klasse und beziehen diese in die Analyse ein.

Ebene	didaktisch beeinflussbare Einflussfaktoren	ausgewählte Fördermöglichkeiten
Buchstaben-, Wort- und Satzerkennung	Lesegenauigkeit (De- und Rekodierung von Wörtern)	intensive Dekodierübungen auf der Wortebene
	Automatisierung der Dekodierfähigkeit	Übungen zur Lautbewusstheit und Lautsynthese
		Lautleseverfahren zur Verbesserung der Leseflüssigkeit und der Fähigkeit zum Sequenzieren von Sätzen
		Wortschatzarbeit
Lokale Kohärenzbildung auf der Satzebene	Bereichsspezifisches Vorwissen	Viellese-Verfahren (freie Lesezeiten)
	Semantische und syntaktische Analysefähigkeit der Sätze	Wortschatzarbeit
Globale Kohärenzbildung	Strategiewissen	Trainieren von Lesestrategien
	Selbsteinschätzung als Leser	
	Motivation	
Bildung von Superstrukturen	Sprach-, Text- und Weltwissen	Literaturunterricht
	Domänenspezifisches Wissen	
Erkennen rhetorischer Strategien	Fähigkeit zur Reflektion der Textstrukturen und rhetorischen Mittel	Produktorientierte Verfahren
Lese- und Lernmotivation	Selbsteinschätzung als Leser	Arbeit mit Selbsteinschätzungsverfahren
	Übernahme von Verantwortung für das eigene Lernen	Lese-Lerntagebuch
	Motivation	Verfahren der Leseanimation

Tabelle 9: Ebenen des Leseprozesses, didaktisch beeinflussbare Einflussfaktoren und ausgewählte Fördermöglichkeiten (vgl. IQB, 2009, S. 69)

Auf hierarchiehöherer Ebene steht u.A. die „Bildung von Superstrukturen“ im Fokus: Unter Superstrukturen sind Regeln und Kategorien zur Ordnung von Texten zu verstehen. Da sich Leserinnen und Leser bei einem neuen Text an ihrem bereits bestehenden Wissen (über die Eigenschaften von Textsorten) orientieren, besitzen Superstrukturen große Bedeutung für den Leseprozess. Die Fähigkeit zur Bildung von Superstrukturen ist wesentlich vom Sprach-, Text-, Welt- und domänenspezifischen Wissen abhängig. Sie kann im Unterricht insbesondere durch die Vermittlung von Wissen über Textsorten gefördert werden. Schülerinnen und Schüler müssen dazu die Möglichkeit erhalten, im Unterricht charakteristische Textstrukturen zu erarbeiten, wobei der Literaturunterricht bereits wesentliche Arbeit leistet. Unterentwickelt scheint bisher allerdings die Vermittlung von Textstrukturcharakteristika bei Sachtexten zu sein, welche sich deutlich von jenen bei narrativen Texten unterscheiden können. Hier ist die Weiterarbeit mit den Testmaterialien und -aufgaben zu empfehlen (vgl. IQB, 2009), indem einem Sachtext aus den Aufgaben der Vergleichsarbeiten ein weiterer Sachtext zum selben oder zu einem ähnlichen Thema zur Seite gestellt wird. Im Unterricht wird dann textvergleichend vorgegangen, indem in Gruppenarbeit der Aufbau beider Texte schematisch gegenübergestellt wird. Die Schülerinnen und Schüler arbeiten so charakteristische Muster heraus.

### Beispiel 2: Von der Testaufgabe zur Lernaufgabe - Förderung der Modellierungskompetenz im Mathematikunterricht

Förderung der Modellierungskompetenz im Mathematikunterricht

Die nachfolgend beschriebene Aufgabe aus der Leitidee „Daten und Zufall“ wurde als Modifikation einer Aufgabe der Bildungsstandards formuliert (Lankes et al., 2005). Sie ermöglicht es darzustellen, wie unter demselben Stimulus durch Öffnung der Aufgabenformate verschiedene Einsatzmöglichkeiten als Test- und als Lernaufgabe im Unterricht konstruiert werden können. Unter den in Kapitel 2.4.2.3 aufgeführten Kompetenzen steht *Modellieren* im Zentrum; die Anforderungsbereiche hängen vom jeweiligen Aufgabenteil ab:

#### Weitspringen, der Springer-Cup

„An der Erich-Hüpf-Schule wird jedes Jahr eine 6. Klasse mit dem „Springer-Cup“ ausgezeichnet. Den Preis erhält die Klasse mit den besten Weitsprungergebnissen. Drei Klassen gehen dieses Mal mit folgenden Weiten ins Rennen (Abbildung 14).“

Sprungweiten (auf 10 cm gerundet)			
Klasse 6a	Klasse 6b	Klasse 6c	
390	390	370	In der Jury gibt es verschiedene Vorschläge, wie der Preis vergeben werden soll:
380	370	380	
380	390	400	
390	400	390	<b>Jury A</b> „Die Sache ist doch klar, den Preis bekommt die Klasse mit dem Spitzenspringer.“
400	390	410	
390	360	410	
390	370	360	<b>Jury B</b> „Mit dem Preis soll doch die gesamte Klasse ausgezeichnet werden und nicht nur der beste Springer. Also gehört der Preis der Klasse, die die meisten guten Springer besitzt.“
380	430	380	
390	400	370	
380	380	360	<b>Jury C</b> „Und was, wenn der Rest grottenschlecht ist? Der ausgeglichensten Klasse gehört der Preis.“
380	370	400	
390	390	390	
390	380	410	
370	410	390	
390	390	390	
380	380	370	
390	390	380	
390	370	410	
380	380	360	
390	370	370	
380	390	380	
380		400	
390		390	

Abbildung 14: Stimulus der Aufgabe „Weitspringen, der Springer-Cup“ in Anlehnung an Lankes et al. (2005, S. 37)

Aufbauend auf diesem Stimulus lassen sich einerseits Testaufgaben zur Kompetenzbestimmung formulieren, wie sie sich etwa bei Vergleichsarbeiten finden, z.B.

1. Bestimme das arithmetische Mittel und den Zentralwert für jede Klasse. Was lässt sich feststellen?
2. Welche Klasse bekäme den Preis, wenn die Vergabe nach der Regelung von A, B oder C vorgenommen werden würde? Begründe jeweils.
3. An welche Klasse sollte deiner Meinung nach der „Springer-Cup“ vergeben werden? Schreibe einen Bericht an die Jury, in dem du deinen Vergabemodus begründest.

Darüber hinaus kann die Aufgabenstellung zu einer Lernaufgabe umformuliert werden, etwa:

*An welche Klasse sollte eurer Meinung nach der Springer-Cup vergeben werden?*

*Schreibt einen Bericht, in dem ihr eure Meinung deutlich macht. Folgende Bereiche sollten darin angesprochen werden: Arithmetischer Mittelwert; Zentralwert; Anfertigen von Säulendiagrammen; Umgang mit Diagrammen und Kennwerten.*

Diese Formulierung bietet Schülerinnen und Schüler die folgenden Möglichkeiten zum Kompetenztraining:

1. Sie bearbeiten eine alltagsnahe mathematische Fragestellung eigenverantwortlich entsprechend ihrer mathematischen Kompetenzen. Dabei spielen die rechnerischen Anforderungen gegenüber der verständnisvollen Entwicklung mathematischer Konzepte eine eher untergeordnete Rolle, auch wenn die Aufgabe durchaus umfangreiche Rechnungen verlangt (Einsatz des Taschenrechners möglich).
2. Eine arbeitsteilige Bewältigung im Rahmen einer Gruppenarbeit ist möglich. Da sich bei den ausgewählten Ergebnissen sowohl durch den arithmetischen Mittelwert als auch den Zentralwert keine klaren Sieger festlegen lassen, sind bei einer Gruppenarbeit unterschiedliche Vorschläge zu erwarten. Die Schülerinnen und Schüler erleben, dass die Entscheidung dann nur normativ in der Diskussion getroffen werden kann.
3. Es besteht die Möglichkeit, eine offene Aufgabenstellung zu formulieren, bei der die Datenaufbereitung von den Schülerinnen und Schüler selbst übernommen wird, die dabei verschiedene Phasen des Modellierungskreislaufs durchlaufen.
4. Eine Modifikation der Aufgabe, beispielsweise auf selbstständig erhobene Daten, ist genauso denkbar wie fächerübergreifendes Arbeiten in Kooperation mit anderen Fächern.

An den beiden vorgestellten Beispielen wird deutlich, dass die Durchführung, Auswertung und Interpretation von Vergleichsarbeiten alleine noch nicht zur Verbesserung des Unterrichts ausreichen, die eigentliche Arbeit vielmehr noch folgt. Sie verdeutlichen allerdings auch: Vergleichsarbeiten können ein Ausgangspunkt für die Entwicklung gehaltvoller Konzepte zum kompetenzorientierten Unterricht sein.

### 2.5.5 Weiterführende Literatur

Die Nutzung von Ergebnissen aus Vergleichsarbeiten für die Unterrichtsentwicklung beschreiben Peek & Döbelstein (2006) sowie Bensen, Büchter & Peek (2006). Zu empfehlen ist das Herausgeberwerk von Kuper & Schneewind (2006) zur Rezeption von Schulleistungsergebnissen, in dem sich unter anderem Groß-Ophoff, Koch, Hosenfeld & Helmke mit der Rezeption der Ergebnissrückmeldungen aus dem VERA-Projekt befassen. Eine systematische Aufarbeitung empirischer Erkenntnisse zur Ergebnisrezeption und -nutzung hat zuletzt Dederich (2011) vorgestellt; ihre Analyse weist auf Optimierungsbedarf bei schulischen Aktivitäten im Anschluss an die Ergebnissrückmeldung hin. Anregungen zur Verwendung der Ergebnissrückmeldung für Prozesse der Unterrichtsentwicklung geben die Webseiten der Ministerien. Die hier vorgestellten Ideen zur Förderung der Lesekompetenz knüpfen an Rosebrock & Nix (2008) an. Der Förderung von Lesekompetenz hat sich das Projekt ProLesen (<http://www leseforum bayern de/index.asp?MNav=6>) verschrieben. Vorschläge zur Entwicklung kompetenzorientierter Mathematikaufgaben bieten z.B. Blum, Drücke-Noe, Hartung & Köller (2006) sowie Büchter & Leuders (2005).

### 2.5.6 Verständnis und Diskussionspunkte

1. Beschreiben Sie bitte, wie die unter 2.5.1 aufgeführten Informationen der Ergebnissrückmeldung genutzt werden können. Für welchen Zweck wird die jeweilige Information zur Verfügung gestellt?
2. Sie erhalten im Rahmen von Vergleichsarbeiten die folgende Rückmeldung über das Abschneiden der Jahrgangsstufe:

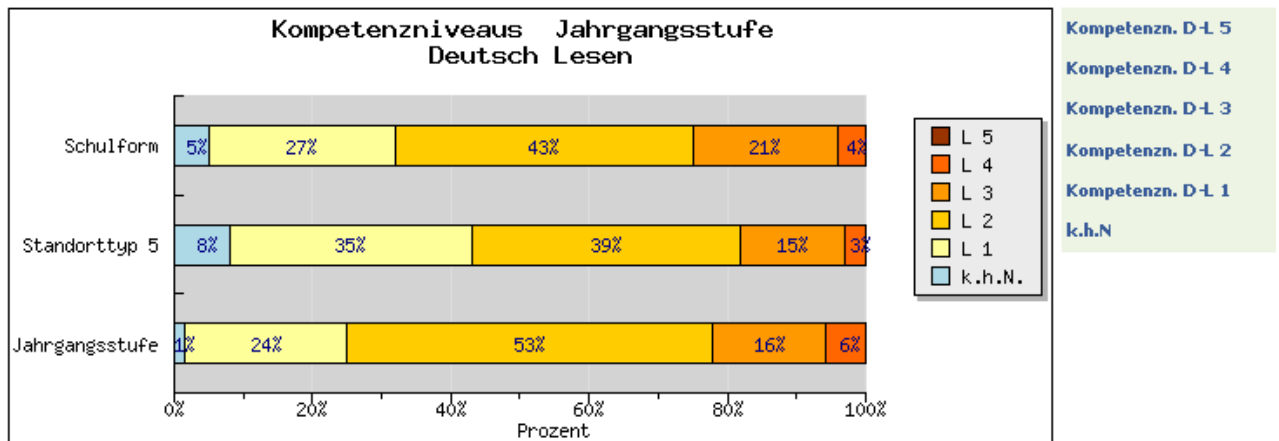


Abbildung 15: Beispiel einer Ergebnissrückmeldung: Verteilung einer Jahrgangsstufe und Referenzverteilungen auf fünf Kompetenzstufen zuzüglich des nicht auswertbaren Bereiches (k.h.N. – „Kein hinreichender Nachweis für das Erreichen eines Kompetenzniveaus“)

Welche Aussagen sind Abbildung 15 zu entnehmen?

3. Im Mittelpunkt der Darstellung dieses letzten Kapitels stand die Zielsetzung, ausgehend von der Ergebnissrückmeldung aus Vergleichsarbeiten Unterrichtsentwicklung zu betreiben. Skizzieren Sie bitte kurz einige Möglichkeiten, Ergebnissrückmeldungen zur Verbesserung der diagnostischen Kompetenzen von Lehrkräften zu nutzen.



## 2.6 Literaturverzeichnis

- Arnold, Karl-Heinz (1999); *Fairneß bei Schulsystemvergleichen; Diagnostische Konsequenzen von Schulleistungsstudien für die unterrichtliche Leistungsbewertung und binnenschulische Evaluation*; Münster, Waxmann
- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001); Lesekompetenz: Testkonzeption und Ergebnisse; In J. Baumert, E. Klieme, M. Neubrand, Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Hrsg.), *PISA 2000: Basiskompetenzen von Schülern im internationalen Vergleich* (S. 69-137); Opladen, Leske & Budrich
- Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J. & Weiß, M. (Hrsg.) (2002); *PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich*; Opladen, Leske + Budrich
- Baumert, J., Trautwein, U. & Artelt, C. (2003); Schulumwelten – institutionelle Bedingungen des Lehrens und Lernens; In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann, & M. Prenzel (Hrsg.); *PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 259- 330); Opladen, Leske + Budrich
- Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J. & Prenzel, M. (Hrsg.) (2003); *PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*; Opladen, Leske + Budrich
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Hrsg.) (2001); *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*; Opladen, Leske + Budrich
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O., & Neubrand, J. (1997); *TIMSS - Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich*; Deskriptive Befunde; Opladen, Leske + Budrich
- Blum, W., Drüke-Noe, C., Hartung, R. & Köller, O. (Hrsg.) (2006); *Bildungsstandards Mathematik: konkret*; Berlin, Cornelsen Skriptor
- Bond, T.G. & Fox, C.M. (2001); *Applying the Rasch Model: fundamental measurement in the human sciences*; Mahwah, NJ, LEA
- Bonsen, M., Büchter, A. & Peek, R. (2006); Datengestützte Schul- und Unterrichtsentwicklung – Bewertungen der Lernstandserhebungen in NRW durch Lehrerinnen und Lehrer; In W. Bos, H.G. Holtappels, H. Pfeiffer, H.-G. Rolff, & R. Schulz-Zander (Hrsg.). *Jahrbuch der Schulentwicklung* (Vol. 14; S. 125-148); Weinheim, Juventa
- Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.-V., Schwippert, K. & Valtin, R. (Hrsg.) (2007); *IGLU 2006 - Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*; Münster, Waxmann
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R., & Walther, G. (Hrsg.) (2003); *Erste Ergebnisse aus IGLU; Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*; Münster, Waxmann
- Bremerich-Vos, A., Granzer, D. & Köller, O. (Hrsg.) (2008); Lernstandsbestimmung im Fach Deutsch. Gute Aufgaben für den Unterricht; Weinheim, Beltz
- Bremerich-Vos, A., Granzer, D., Behrens, U. & Köller, O. (Hrsg.) (2009); *Bildungsstandards für die Grundschule: Deutsch konkret: Aufgabenbeispiele – Unterrichts Anregungen - Fortbildungsideen*; Berlin, Cornelsen Skriptor
- Büchter, A. & Leuders, T. (2005). *Mathematikaufgaben selbst entwickeln. Lernen fördern – Leistung prüfen*. Berlin, Cornelsen Skriptor
- Burkard, C. & Peek, R. (2004); Anforderungen an zentrale Lernstandserhebungen; Ein Werkstattbericht aus Nordrhein-Westfalen; *Pädagogik*, 6, 24-27
- Dedering, K. (2011). Hat Feedback eine positive Wirkung? Zur Verarbeitung extern erhobener Leistungsdaten in Schulen; *Unterrichtswissenschaft*, 39, 63-83
- Dobbelstein, P. & Peek, R. (2007); Einleitung: Lernstandserhebungen als Beitrag zu einer empiriegestützten Unterrichtsentwicklung; In MSW (Hrsg.), *Lernstandserhebungen Mathematik in Nordrhein-Westfalen; Impulse zum Umgang mit zentralen Tests* (S. 7-13); Stuttgart, Klett
- DESI-Konsortium (Hrsg.) (2008); *Unterricht und Kompetenzerwerb in Deutsch und Englisch; Ergebnisse der DESI-Studie*; Weinheim, Beltz
- Eikenbusch, G. & Leuders, T. (Hrsg.) (2004); *Lehrer-Kursbuch Statistik*; Berlin, Cornelsen Skriptor
- Europarat (2001); *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*; Berlin, Langenscheidt
- Fleischer, J., Wirth, J. & Leutner, D. (2007); Testmethodische Grundlagen der Lernstandserhebungen NRW: Erfassung von Schülerkompetenzen für Vergleiche mit kriterialen und sozialen Bezugsnormen; In MSW (Hrsg.), *Lernstandserhebungen Mathematik in Nordrhein-Westfalen; Impulse zum Umgang mit zentralen Tests* (S. 91-113); Stuttgart, Klett
- Freudenthal, H. (1977); *Mathematik als pädagogische Aufgabe (Bd. 1, Bd. 2)*; Stuttgart, Ernst Klett Verlag
- Freudenthal, H. (1983); *Didactical phenomenology of mathematical structures*; Dordrecht, Reidel
- Groß Ophoff, J., Koch, U., Hosenfeld, I. & Helmke, A. (2006); Ergebnissrückmeldungen und ihre Rezeption im Projekt VERA; In H. Kuper & J. Schneewind (Hrsg.), *Rückmeldung und Rezeption von Forschungsergebnissen - Zur Verwendung wissenschaftlichen Wissens im Bildungssystem* (S. 19-40); Münster, Waxmann
- Helmke, A. & Hosenfeld, I. (2004); Vergleichsarbeiten - Standards - Kompetenzstufen: Begriffliche Klärung und Perspektiven; In R. S. Jäger, A. Frey & M. Wosnitza (Hrsg.); *Lernprozesse, Lernumgebungen und Lern Diagnostik; Wissenschaftliche Beiträge zum Lernen im 21. Jahrhundert* (S. 56-75); Landau, Verlag Empirische Pädagogik
- Helmke, A. & Hosenfeld, I. (2003a); Vergleichsarbeiten (VERA): eine Standortbestimmung zur Sicherung schulischer Kompetenzen - Teil 1: Grundlagen, Ziele, Realisierung; *SchulVerwaltung NRW*, 4, 107-110
- Helmke, A. & Hosenfeld, I. (2003b); Vergleichsarbeiten (VERA): eine Standortbestimmung zur Sicherung schulischer Kompetenzen - Teil 2: Nutzung für Qualitätssicherung und Verbesserung der Unterrichtsqualität; *SchulVerwaltung NRW*, 5, 143-145
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004); Vergleichsarbeiten als Werkzeug für die Verbesserung der diagnostischen Kompetenz von Lehrkräften; In R. Arnold & C. Giese (Hrsg.); *Schulleitung und Schulentwicklung* (S. 119-144); Hohengehren, Schneider



- Helmke, A. & Schrader, F. W. (2001); Determinanten der Schulleistung; In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (2. Aufl., S. 81-91); Weinheim, Beltz
- IQB (ohne Jahr); *Deutsch Sekundarstufe I*; Verfügbar unter: [https://www.iqb.hu-berlin.de/arbbereiche/projekte/?pg=p\\_36](https://www.iqb.hu-berlin.de/arbbereiche/projekte/?pg=p_36) [24.07.2009]
- IQB (2009); *Deutsch: VERA 8 Handreichung 2009 - Testheft II*. Verfügbar unter: [http://www.iqb.hu-berlin.de/vera2?reg=r\\_6](http://www.iqb.hu-berlin.de/vera2?reg=r_6) [24.07.2009]
- IQB (2010); *Vergleichsarbeiten 2010. 3. Jahrgangsstufe (VERA-3). Mathematik – Didaktische Handreichung*. Verfügbar unter: [http://www.standardsicherung.schulministerium.nrw.de/vera3/upload/download/mat\\_10-11/VERA\\_M\\_Did\\_Hand\\_Mathematik.pdf](http://www.standardsicherung.schulministerium.nrw.de/vera3/upload/download/mat_10-11/VERA_M_Did_Hand_Mathematik.pdf)
- Isaac, K. (2011); Neues Standorttypenkonzept. Faire Vergleiche bei Lernstandserhebungen. *Schule NRW 06/11. Amtsblatt des Ministeriums für Schule und Weiterbildung*, 300-301
- Isaac, K., Metzeld, D. & Eichler, W. (2009); Bewusster Umgang mit Sprache – Sprache und Sprachgebrauch untersuchen; *Grundschulunterricht*, 56 (2), 28-31
- Klauer, K.J. (1987); *Kriteriumsorientierte Tests*; Göttingen, Hogrefe
- Kliemann S. (Hrsg.) (2010); *Diagnostizieren und Fördern - Kompetenzen erkennen, unterstützen und erweitern. Beispiele und Anregungen für die Jahrgänge 1 bis 4*; Berlin, Cornelsen Verlag Scriptor
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.E. & Vollmer, J. (2003); Zur Entwicklung nationaler Bildungsstandards; Eine Expertise; Berlin, BMBF
- Klieme, E., Hartig, J. & Rauch, D. (2008); The concept of competence in educational contexts; In J. Hartig, E. Klieme, D. Leutner (Hrsg.); *Assessment of competencies in educational contexts* (S. 3-22); Göttingen, Hogrefe & Huber
- Klieme, E. & Leutner, D. (2006); Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen; Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG; *Zeitschrift für Pädagogik*, 52, 876-903
- KMK (2003a); *Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss: Beschluss vom 04. 12. 2003*; München, Luchterhand
- KMK (2003b); *Bildungsstandards für die erste Fremdsprache (Englisch / Französisch) für den Mittleren Schulabschluss: Beschluss vom 04. 12. 2003*; München, Luchterhand
- KMK (2003c); *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss: Beschluss vom 04. 12. 2003*; München, Luchterhand
- KMK (2004a); *Bildungsstandards im Fach Deutsch für den Primarbereich: Beschluss vom 15. 10. 2004*; München, Luchterhand
- KMK (2004b); *Bildungsstandards im Fach Mathematik für den Primarbereich: Beschluss vom 15. 10. 2004*; München, Luchterhand
- KMK (2004c); *Bildungsstandards im Fach Deutsch für den Hauptschulabschluss: Beschluss vom 15. 10. 2004*; München, Luchterhand
- KMK (2004d); *Bildungsstandards für die erste Fremdsprache (Englisch / Französisch) für den Hauptschulabschluss: Beschluss vom 15. 10. 2004*; München, Luchterhand
- KMK (2004e); *Bildungsstandards im Fach Mathematik für den Hauptschulabschluss: Beschluss vom 15. 10. 2004*; München, Luchterhand
- KMK & for.mat (ohne Jahr); *Fortbildungskonzepte und -materialien zur kompetenz- bzw. standardbasierten Unterrichtsentwicklung*. Verfügbar unter: <http://www.kmk-format.de/> [22.07.2010].
- KMK & IQB (2006); *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*; München, Luchterhand
- Köller, O. & Baumert, J. (2002); Entwicklung von Schulleistungen; In R. Oerter & L. Montada (Hrsg.); *Entwicklungspsychologie* (5. Aufl., S. 756 - 786); Weinheim, Beltz/PVU
- Köller, O. (2008); Bildungsstandards in Deutschland: Implikationen für die Qualitätssicherung und Unterrichtsqualität; *Zeitschrift für Erziehungswissenschaft, Sonderheft 9*, 47-59
- Kuper, H. & Schneewind, J. (2006); *Rückmeldung und Rezeption von Forschungsergebnissen; Zur Verwendung wissenschaftlichen Wissens im Bildungsbereich*; Münster, Waxmann
- Lankes, E.-M., Lorenzen, H., Petersen, C., von Urban, S. & Zielinski, D. (2005). *Kompetenzorientierter Mathematikunterricht. Anregungen für die Arbeit mit den Bildungsstandards zum Hauptschulabschluss und mittleren Abschluss (Sekundarstufe I)*. Kronshagen: Institut für Qualitätsentwicklung an Schulen.
- Lehmann, R. H. & Peek, R. (1997); *Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen; Bericht über die Untersuchung im September 1996 (unveröffentlichter Forschungsbericht)*; Hamburg
- Leutner, D. (2010); Pädagogisch-psychologische Diagnostik; In D. H. Rost (Hrsg.); *Handwörterbuch Pädagogische Psychologie* (4. überarbeitete Aufl., S. 624-635); Weinheim, PVU
- Leutner, D., Fleischer, J., Spoden, C. & Wirth, J. (2007); Landesweite Lernstandserhebungen zwischen Bildungsmonitoring und Individualdiagnostik; *Zeitschrift für Erziehungswissenschaft, Sonderheft 8*, 149-167
- Lienert, G.A. & von Eye, A. (1998); *Erziehungswissenschaftliche Statistik; Eine elementare Einführung für pädagogische Berufe*; Weinheim, Beltz
- Möller, J. & Schiefele, U. (2004); Motivationale Grundlagen der Lesekompetenz; In U. Schiefele, C. Artelt, W. Schneider, & P. Stanat (Hrsg.); *Entwicklung, Bedingungen und Förderung der Lesekompetenz: Vertiefende Analysen der PISA-2000-Daten* (S. 101-124); Wiesbaden, Verlag für Sozialwissenschaften
- Möller, G., Pallack, A. & Fleischer, J. (2007); Da schau hin: Was Lehrerinnen und Lehrer aus Lernstandserhebungen über ihre schwachen Schülerinnen und Schüler erfahren können; In A. Peter-Koop & A. Bikner-Ahsbahr (Hrsg.); *Mathematische Bildung - Mathematische Leistung; Festschrift für Michael Neubrand zum 60. Geburtstag* (S. 97-113). Hildesheim, Franzbecker

- Nachtigall, C., Kröhne, U., Enders, U. & Steyer, R. (2008); Considering the influence of context variables on students competencies; In J. Hartig, E. Klieme & D. Leutner (Hrsg.); *Assessment of Competencies in Educational Contexts* (S. 315-335); Göttingen, Hogrefe & Huber
- Nachtigall, C., Kröhne, U. & Müller, M. (2005); *Ergänzungen zum Ergebnisbericht der Kompetenztests 2005*; Verfügbar unter: [http://www.kompetenztest.de/download/kt05/Muster\\_Ergaenzg\\_MK06.pdf](http://www.kompetenztest.de/download/kt05/Muster_Ergaenzg_MK06.pdf) [24.07.2009]
- Nachtigall, C. & Jantowski, A. (2004); Die Thüringer Kompetenztests; *Neue Praxis der Schulleitung, Thüringen*, 73, 1-14
- Netzwerk Empiriegestützte Schulentwicklung (2006); *Positionspapier zu: Zentrale standardisierte Lernstandserhebungen*; Verfügbar unter: [http://www.iqb.hu-berlin.de/bista/dateien/EMSE\\_Positionsp.pdf](http://www.iqb.hu-berlin.de/bista/dateien/EMSE_Positionsp.pdf) [24.07.2009]
- Peek, R. (2004); Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik (Qua-SUM) – Klassenbezogene Ergebnisrückmeldungen und ihre Rezeption in Brandenburger Schulen; *Empirische Pädagogik*, 18 (Themenheft), 82-114
- Peek, R. & Döbelstein, P. (2006); Zielsetzung: Ergebnisorientierte Schul- und Unterrichtsentwicklung; Potenziale und Grenzen der nordrhein-westfälischen Lernstandserhebungen; In Böttcher, W., Holtappels, H. G. & Brohm, M. (Hrsg.); *Evaluation im Bildungswesen; Eine Einführung in Grundlagen und Praxisbeispiele* (Grundlagentexte Pädagogik, S. 177-194); Weinheim und München, Juventa
- Projekt VERA (2008). *Didaktische Erläuterungen „Leseverständnis“ und „Sprache und Sprachgebrauch untersuchen“*. Verfügbar unter: [http://139.14.28.6/verapub/fileadmin/downloads/2008/VERA\\_D\\_didakt\\_Erlaeut\\_2008.pdf](http://139.14.28.6/verapub/fileadmin/downloads/2008/VERA_D_didakt_Erlaeut_2008.pdf) [28.01.2010]
- Projekt VERA (2009); *Korrekturanweisungen für die Deutschaufgaben 2009*. Verfügbar unter: [http://139.14.28.6/verapub/fileadmin/downloads/2009/VERA\\_D\\_Korrekturanweisung\\_2009.pdf](http://139.14.28.6/verapub/fileadmin/downloads/2009/VERA_D_Korrekturanweisung_2009.pdf) [28.01.2010]
- Ramm, G., Prenzel, M., Heidemeier, H. & Walter, O. (2004); Soziokulturelle Herkunft: Migration; In Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolff, H.-G., Rost, J. & Schiefele, U. (Hrsg.); *Der Bildungsstand der Jugendlichen in Deutschland - Ergebnisse des zweiten internationalen Vergleichs* (S. 254-272); Münster, Waxmann
- Rasch, G. (1960); *Probabilistic models for some intelligence or attainment tests*; Copenhagen, Nielsen & Lydiche (2<sup>nd</sup> Edition Chicago University of Chicago Press, 1980)
- Reiss, K. & Winkelmann, H. (2009); Kompetenzstufenmodelle für das Fach Mathematik im Primarbereich; In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik Leistungsmessung in der Grundschule* (S. 120-141); Weinheim, Beltz
- Richter, T. & Christmann, U. (2002); Lesekompetenz: Prozessebenen und interindividuelle Unterschiede. In N. Groeben & B. Hurrelmann (Hrsg.), *Lesekompetenz: Bedingungen, Dimensionen, Funktionen* (S. 25-59); Weinheim, Juventa
- Rosebrock, C. & Nix, D. (2008); *Grundlagen der Lesedidaktik und der systematischen Leseförderung*. Baltmannsweiler, Hohengehren
- Rost, J. (2004); *Lehrbuch Testtheorie - Testkonstruktion*; Bern, Huber
- Ruep, M & Keller, G. (2007); *Schulevaluation*; Frankfurt, Peter Lang
- Rupp, A. A., Vock, M., Harsch, C. & Köller, O. (2008); *Developing standards-based assessment items for English as a first foreign language – Context, processes, and outcomes in Germany*; Münster, Waxmann
- Schnotz, W. & Dutke, S. (2004); Kognitionspsychologische Grundlagen der Lesekompetenz: Mehrebenenverarbeitung anhand multipler Informationsquellen; In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.); *Struktur, Entwicklung und Förderung von Lesekompetenz; Vertiefende Analysen im Rahmen von PISA 2000* (S. 61-99); Wiesbaden, VS Verlag für Sozialwissenschaften
- Schrader, F.-W. (2001); Diagnostische Kompetenz von Eltern und Lehrern; In D.H. Rost (Hrsg.); *Handwörterbuch Pädagogische Psychologie* (2. überarb. u. erw. Aufl., S. 91-96); Weinheim, Beltz
- Schrader, F.-W. & Helmke, A. (2001); Alltägliche Leistungsbeurteilung durch Lehrer; In F. E. Weinert (Hrsg.); *Leistungsmessungen in Schulen* (S. 45-58); Weinheim, Beltz
- Schräpler, J.-P. (2009); Verwendung von SGB II-Dichten als Raumindikator für die Sozialberichterstattung am Beispiel der „sozialen Belastung“ von Schulstandorten in NRW – ein Kernel-Density-Ansatz. *Statistische Analysen und Studien Nordrhein-Westfalen*, 57, 3-28
- Spinath, B. (2005); Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer/innen und das Konstrukt der diagnostischen Kompetenz; *Zeitschrift für Pädagogische Psychologie*, 19, 85-95
- Süß, H. M. (2001); Prädiktive Validität der Intelligenz im schulischen und außerschulischen Bereich; In E. Stern & J. Guthke (Hrsg.); *Perspektiven der Intelligenzforschung* (S. 109-135); Lengerich, Pabst Science Publishers
- Tenorth, H.-E. (2001); Englisch: Ein Kerncurriculum, seine Notwendigkeit und seine Gestalt – Zusammenfassung; In H.-E. Tenorth, (Hrsg.); *Kerncurriculum Oberstufe; Mathematik – Deutsch – Englisch; Expertisen im Auftrag der Ständigen Konferenz der Kultusminister* (S. 156-161); Weinheim, Beltz
- Tesch, B., Leupold, E., Köller, O. (Hrsg.) (2008); *Bildungsstandards Französisch: konkret*; Berlin, Cornelsen Scriptor
- Walther, G., van den Heuvel-Panhuizen, M., Granzer, D. & Köller, O. (2007); *Bildungsstandards für die Grundschule: Mathematik konkret*; Berlin, Cornelsen Skriptor
- Weinert, F. E. (2001); Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit; In Weinert, F. E. (Hrsg.); *Leistungsmessungen in Schulen* (S. 17-31); Weinheim, Beltz
- Wirtz, M. & Caspar, F. (2002); *Beurteilerübereinstimmung und Beurteilerreliabilität*; Göttingen, Hogrefe
- Ziener, G. (2006); *Bildungsstandards in der Praxis. Kompetenzorientiert unterrichten*; Seelze, Velber